

Datenanalyse zu potentiellen Unterschieden im Trinkwasserbedarf verschiedener Siedlungsgebiete anhand siedlungsstruktureller Kennzahlen unter Verwendung eines automatisierten FME (Feature Manipulation Engine)-Prozesses

Data analysis on potential differences in the drinking water demand of different settlement areas by means of structural characteristics using an automated FME (feature manipulation engine) process

Masterarbeit von Frederik S. Kip

Datum: 19. Mai 2021, Tag der Einreichung: 28. Februar 2018

1. Gutachten: Prof. Dipl.-Ing. Dr. nat. techn. Wilhelm Urban (IWAR)

2. Gutachten: Dr.-Ing. Alexander Sonnenburg (IWAR)

Darmstadt



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Fachbereich Bau- und
Umweltingenieurwissen-
schaften

Institut IWAR

Fachgebiet
Wasserversorgung und
Grundwasserschutz

Datenanalyse zu potentiellen Unterschieden im
Trinkwasserbedarf verschiedener Siedlungsgebiete
anhand siedlungsstruktureller Kennzahlen unter
Verwendung eines automatisierten FME (Feature
Manipulation Engine)-Prozesses

Data analysis on potential differences in the
drinking water demand of different settlement areas
by means of structural characteristics using an
automated FME (feature manipulation engine)
process

Masterarbeit von Frederik S. Kip

1. Gutachten: Prof. Dipl.-Ing. Dr. nat. techn. Wilhelm
Urban (IWAR)

2. Gutachten: Dr.-Ing. Alexander Sonnenburg
(IWAR)

Datum: 19. Mai 2021

Tag der Einreichung: 28. Februar 2018

Darmstadt

Bitte zitieren Sie dieses Dokument als:
URN: urn:nbn:de:tuda-tuprints-90634
URL: <http://tuprints.ulb.tu-darmstadt.de/9063>

Dieses Dokument wird bereitgestellt von tuprints,
E-Publishing-Service der TU Darmstadt
<http://tuprints.ulb.tu-darmstadt.de>
tuprints@ulb.tu-darmstadt.de

Die Veröffentlichung steht unter folgender Creative Commons Lizenz:
Namensnennung – Nicht kommerziell – Weitergabe unter gleichen Bedingungen 4.0 International
<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Erklärung zur Master-Thesis

Hiermit versichere ich, die vorliegende Master-Thesis ohne Hilfe Dritter nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die aus Quellen entnommen wurden, sind als solche kenntlich gemacht. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Inhaltsverzeichnis

Abbildungsverzeichnis	iii
Tabellenverzeichnis	v
Abkürzungsverzeichnis	vi
1. Einleitung	1
2. Theoretische Grundlagen und Methoden	3
2.1. Wasserbedarfsprognose	4
2.1.1. Aktuelle Forschungen	7
2.2. Siedlungsstruktur	16
2.2.1. Siedlungsformen, Klassifizierung	16
2.2.2. Siedlungsstrukturelle Kennzahlen	18
2.2.3. Bodenrichtwerte	23
2.2.4. Raumordnung	24
2.3. Geoinformationssysteme	27
2.3.1. Verwendete Software	33
2.4. Statistische Analyse	36
2.4.1. Univariate Häufigkeitsverteilungen	37
2.4.2. Regression und Korrelation	41
2.4.3. Clusteranalyse	42
2.4.4. Verwendete Software	44
3. Datengrundlage	46
3.1. Bestandsdatenauszug	46
3.2. 3D-Gebäudedaten	52
3.3. Bodenrichtwerte	55
3.4. Zensus-Daten	57
3.5. Sonstige Daten	58
4. Datenaufbereitung und -analyse	59
4.1. Datenreduktion	60
4.2. Datensynthese	63
4.2.1. Synthese der Flurstücksdaten	63
4.2.2. Synthese der Gebäudedaten	69
4.2.3. <i>GF</i> - Verheiraten der Daten - Gebäude und Flurstücke	73
4.3. Datenanalyse	76
5. Ergebnisse und Diskussion	79
5.1. Ergebnis der Literaturrecherche	79
5.2. Diskussion über die Datengrundlage	81
5.3. Ergebnisse und Diskussion der Datenaufbereitung	82



5.4. Ergebnisse und Diskussion der Datenanalyse	93
5.5. Fazit und Kritik	100
6. Ausblick	102
Literatur	104
A. Ergebnisse der Datenaufbereitung	109
B. Ergebnisse der Datenanalyse	124

Abbildungsverzeichnis

2.1. Beispielhafte Darstellung eines Flurstücks mit Gebäuden und Geometriedaten	18
4.1. FME-Workspace <i>Z0</i> - Clippen von Zensus-Daten mit Puffer	60
4.2. FME-Workspace zur Datensynthese	64
4.3. FME-Workspace <i>F0</i> - Clippen der Flurstücke	65
4.4. FME-Workspace <i>F1</i> - Aggregation der Flurstücke	65
4.5. FME-Workspace <i>F2</i> - Projizieren der Zensus-Daten auf Flurstücke	67
4.6. FME-Workspace <i>F3</i> - Verheiraten Flurstücke und Nutzungsart	68
4.7. FME-Workspace <i>F4</i> - Projizieren der Bodenrichtwerte auf Flurstücke	69
4.8. FME-Workspace <i>G0</i> - Clippen der 3D-Gebäude und Reduktion der XML-Ebenen	71
4.9. FME-Workspace <i>G1</i> - Reduktion der 3D-Daten auf 2D	72
4.10. FME-Workspace <i>G2</i> - Verheiraten von Gebäudedaten	73
4.11. FME-Workspace <i>GF</i> - Verheiraten der Daten - Gebäude und Flurstücke	75
4.12. FME Custom Transformer LoopGebaeude	75
4.13. FME Workspace <i>C1</i> - Clusteranalyse	77
4.14. R-Script zur Clusteranalyse	78
5.1. Ergebnis <i>Z0</i> Reduktion der Zensus-Daten auf Zielgebiet, hier: Gemeinde Alpen	82
5.2. Ergebnis <i>F0</i> Clippen der Flurstücke, hier: Gemeinde Alpen	83
5.3. Ergebnis <i>F1</i> Aggregation der Flurstücke, hier: Gemeinde Alpen	84
5.4. Ergebnis <i>F2</i> Projizieren der Zensus-Daten auf Flurstücke, hier: Gemeinde Alpen . . .	84
5.5. Ergebnis <i>F3</i> Verheiraten Flurstücke und Nutzungsart, hier: Gemeinde Alpen	85
5.6. Ergebnis <i>F4</i> Projizieren der Bodenrichtwerte auf Flurstücke, hier: Gemeinde Alpen . .	85
5.7. Ergebnis <i>G0</i> Clippen der 3D-Gebäude und Reduktion der XML-Ebenen, hier: Gemeinde Alpen	86
5.8. Ergebnis <i>G2</i> Verheiraten von Gebäudedaten, hier: Gemeinde Alpen	87
5.9. Ergebnis <i>GF</i> Verheiraten der Daten - Grundflächenzahl, hier: Gemeinde Alpen	88
5.10. Ergebnis <i>GF</i> Verheiraten der Daten - Geschossflächenzahl, hier: Gemeinde Alpen . . .	88
5.11. Ergebnis <i>GF</i> Verheiraten der Daten - Baumassenzahl, hier: Gemeinde Alpen	89
5.12. Ergebnis <i>GF</i> Verheiraten der Daten - Kompaktheit, hier: Gemeinde Alpen	89
5.13. Beispiel Berechnung Außenfläche: Gebäude aus zwei Teilen	90
5.14. Kern-Dichte-Schätzer, hier: Gemeinde Alpen	94
5.15. Ergebnis <i>GF</i> Clusteranalyse, hier: Gemeinde Alpen	95
5.16. Boxplots zur Visualisierung der Ergebnisse, hier: Gemeinde Alpen	98
A.1. Ergebnis <i>Z0</i> Reduktion der Zensus-Daten auf Zielgebiet, hier: Stadt Lemgo	109
A.2. Ergebnis <i>F0</i> Clippen der Flurstücke, hier: Stadt Lemgo	110
A.3. Ergebnis <i>F1</i> Aggregation der Flurstücke, hier: Stadt Lemgo	110
A.4. Ergebnis <i>F2</i> Projizieren der Zensus-Daten auf Flurstücke, hier: Stadt Lemgo	111
A.5. Ergebnis <i>F3</i> Verheiraten Flurstücke und Nutzungsart, hier: Stadt Lemgo	111
A.6. Ergebnis <i>F4</i> Projizieren der Bodenrichtwerte auf Flurstücke, hier: Stadt Lemgo	112
A.7. Ergebnis <i>G0</i> Clippen der 3D-Gebäude und Reduktion der XML-Ebenen, hier: Stadt Lemgo	113
A.8. Ergebnis <i>G2</i> Verheiraten von Gebäudedaten, hier: Stadt Lemgo	114
A.9. Ergebnis <i>GF</i> Verheiraten der Daten - Grundflächenzahl, hier: Stadt Lemgo	114
A.10. Ergebnis <i>GF</i> Verheiraten der Daten - Geschossflächenzahl, hier: Stadt Lemgo	115

A.11.Ergebnis <i>GF</i> Verheiraten der Daten - Baumassenzahl, hier: Stadt Lemgo	115
A.12.Ergebnis <i>GF</i> Verheiraten der Daten - Kompaktheit, hier: Stadt Lemgo	116
A.13.Ergebnis <i>Z0</i> Reduktion der Zensus-Daten auf Zielgebiet, hier: Stadt Solingen	116
A.14.Ergebnis <i>F0</i> Clippen der Flurstücke, hier: Stadt Solingen	117
A.15.Ergebnis <i>F1</i> Aggregation der Flurstücke, hier: Stadt Solingen	117
A.16.Ergebnis <i>F2</i> Projizieren der Zensus-Daten auf Flurstücke, hier: Stadt Solingen	118
A.17.Ergebnis <i>F3</i> Verheiraten Flurstücke und Nutzungsart, hier: Stadt Solingen	118
A.18.Ergebnis <i>F4</i> Projizieren der Bodenrichtwerte auf Flurstücke, hier: Stadt Solingen	119
A.19.Ergebnis <i>G0</i> Clippen der 3D-Gebäude und Reduktion der XML-Ebenen, hier: Stadt Solingen	120
A.20.Ergebnis <i>G2</i> Verheiraten von Gebäudedaten, hier: Stadt Solingen	121
A.21.Ergebnis <i>GF</i> Verheiraten der Daten - Grundflächenzahl, hier: Stadt Solingen	121
A.22.Ergebnis <i>GF</i> Verheiraten der Daten - Geschossflächenzahl, hier: Stadt Solingen	122
A.23.Ergebnis <i>GF</i> Verheiraten der Daten - Baumassenzahl, hier: Stadt Solingen	122
A.24.Ergebnis <i>GF</i> Verheiraten der Daten - Kompaktheit, hier: Stadt Solingen	123
B.1. Ergebnis <i>GF</i> Clusteranalyse, hier: Stadt Lemgo	124
B.2. Boxplots zur Visualisierung der Ergebnisse, hier: Stadt Lemgo	125
B.3. Ergebnis <i>GF</i> Clusteranalyse, hier: Solingen Mitte	127
B.4. Boxplots zur Visualisierung der Ergebnisse, hier: Solingen Mitte	128

Tabellenverzeichnis

2.1. Institutionen als Lieferanten für Daten nach Kluge, Deffner et al. 2008, S. 767	11
2.2. Klassifizierung durch Einwohnerzahl nach Korda (2005, S. 81)	16
3.1. Attribute und Relationen der Objektart AX_Flurstueck, Kennung: 11001, Auszug aus Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) 2008, S. 43 ff.	47
3.2. Attribute und Relationen der Objektart AX_Flurstueck_Kerndaten, Kennung: 11004, Auszug aus Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) 2008, S. 55 ff.	47
3.3. Attribute und Relationen der Objektart AX_Gemeindekennzeichen, Kennung: 73014, Auszug aus Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bun- desrepublik Deutschland (AdV) 2008, S. 456 f.	48
3.4. Attribute und Relationen der Objektart AX_Gebaeude Auszug aus Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) 2008, S. 190 ff.	49
3.5. Ermäßigungsfaktoren nach Objektanzahl nach AdV 2016, S. 3	53
3.6. Verfügbarkeiten der 3D-Gebäudedaten in den einzelnen Bundesländern, Stand De- zember 2017	54
3.7. Verfügbarkeiten der Bodenrichtwerte in den einzelnen Bundesländern, Stand August 2017	56
5.1. Manuelle Berechnung der Außenfläche eines Gebäudes	91
5.2. Statistische Kennzahlen der Cluster, hier: Gemeinde Alpen	97
B.1. Statistische Kennzahlen der Cluster, hier: Stadt Lemgo	126
B.2. Statistische Kennzahlen der Cluster, hier: Solingen Mitte	129

Abkürzungsverzeichnis

AAA	AFIS-ALKIS-ATKIS	46
AFIS	Amtliches Festpunkt-Informationssystem	50
ALKIS	Amtliches Topographisch-Kartographisches Informationssystem	46
ATKIS	Amtliches Liegenschaftskataster-Informationssystem	50
BauGB	Baugesetzbuch	22
BauNVO	Baunutzungsverordnung	18
BRW	Bodenrichtwert	23
BMZ	Baumassenzahl	18
DIN	Deutsche Industrienorm	5
DVGW	Deutscher Verein des Gas- und Wasserfaches e. V.	13
GFZ	Geschossflächenzahl	18
GML	Geography Markup Language	51
GRZ	Grundflächenzahl	18
GIS	Geoinformationssystem	11
ImmoWertV	Immobilienwertermittlungsverordnung	23
IWW	Rheinisch-Westfälisches Institut für Wasser - Beratungs- und Entwicklungsgesellschaft mbH	1
LoD	Level of Detail	52
NAS	Normbasierte Austauschschnittstelle	51
UML	Unified Modeling Language	50
WHG	Wasserhaushaltsgesetz	4
WBP	Wasserbedarfsprognose	10
XML	Extensible Markup Language	51

1. Einleitung

Infrastrukturen für die Versorgung mit Trinkwasser werden in der Regel für einen Zeitraum von 40 bis 60 Jahren^{1.1} geplant und ausgelegt. Dabei ist darauf zu achten, dass diese zweckmäßig dimensioniert sind. So kann eine Überdimensionierung der Versorgungsleitungen (im Durchmesser) zu Stagnation und dem Risiko einer Verkeimung des Trinkwassers führen; eine zu geringe Auslegung des Rohrleitungsquerschnitts birgt das Risiko einer Unterversorgung. Die Versorgungssituation, d. h. die Verfügbarkeit von Trinkwasser in ausreichender Menge und Qualität ist vom Gesetzgeber vorgeschrieben^{1.2}.

Wasserbedarfsprognosen, die zum Teil sehr detailliert ausfallen können, dienen dazu, den potentiellen Wasserbedarf abzuschätzen. Ein wesentlicher Faktor für den Wasserbedarf eines Versorgungsgebietes ist die Zahl der angeschlossenen Einwohner. Diese können einzeln erfasst werden, was einen erheblichen Aufwand bedeutet und nur zu einem bestimmten Zeitpunkt gilt, oder für ein größeres Gebiet abgeschätzt werden. Diese Abschätzung kann durch eine statistische Auswertung einer größeren Datengrundlage erfolgen. Da in der Bundesrepublik Deutschland städtebauliche Aktivitäten bestimmten Regeln und Gesetzen unterliegen, liegt es nahe, dass sich daraus Ähnlichkeiten in der Einwohnerdichte und Siedlungsstruktur ableiten lassen können. Dabei lautet die Fragestellung:

Gibt es einen signifikanten statistischen Zusammenhang zwischen der Zahl der angeschlossenen Einwohner und der Struktur eines Siedlungsgebietes?

Es soll eine Methode gefunden werden, die es ermöglicht, die Anzahl der Verbraucher als ein Einflussfaktor auf den Wasserbedarf abzuschätzen, so dass diese in eine integrierte Wasserbedarfsprognose einfließen kann. Die Ergebnisse dieser Arbeit sollen in das Arbeitspaket (AP) 3 „Erforschung und hydraulische Modellierung/Simulation von Wirkmechanismen zwischen Einflussfaktoren und Trinkwasserverteilung inkl. hochauflösender hydraulischer Simulation des Verbrauchsverhaltens und Erstellung von neuen Modellierungsansätzen“ des *Innogy* Forschungsvorhaben „Verbesserte Zielnetzplanung und nachhaltige Rehabilitationsstrategie unter Berücksichtigung von Planungsunsicherheiten - Flexnetzplanung“ am Standort Rhein-Main des IWW^{1.3} einfließen. Die Laufzeit dieses Projektes ist von Februar 2017 bis August 2019 angesetzt.

Für den Vergleich von verschiedenen Siedlungstypen müssen zuerst Kenndaten aus einer Datengrundlage synthetisiert werden. Eine solche Datengrundlage steht mittlerweile in Form von Geodaten offen

^{1.1}Wert nach AfA- Tabelle für Versorgungsleitungen, vgl. Fritsch et al. (2014, S. 848) oder Karger, Cord-Landwehr und Hoffmann (2008, S. 3)

^{1.2}Stichwort *Versorgungssicherheit*, vgl. Fritsch et al. (2014, S. 4 f.)

^{1.3}Rheinisch-Westfälisches Institut für Wasser - Beratungs- und Entwicklungsgesellschaft mbH (IWW)

und frei zur Verfügung. Sie besteht aus verschiedenen Katasterdaten, dreidimensionalen Gebäudedaten sowie Daten zur Verteilung von Einwohnerdichten und Grundstückspreisen. Aus den beiden zuerst genannten lassen sich durch Datensynthese siedlungsspezifische Kenndaten zur Struktur und Zusammensetzung eines Siedlungsgebietes isolieren. Durch Klassenbildungsverfahren (Clusteranalysen) lassen sich Gebiete mit ähnlichen Merkmalen zusammenfassen. Diese Klassen können mit einander verglichen werden, so dass sich ihre Eigenschaften bezüglich der Einwohnerdichten und somit auch des potentiellen Trinkwasserbedarfs auf andere Gebiete, z. B. Neubaugebiete oder Transformationsgebiete, übertragen lassen. Dadurch dass in dieser Arbeit siedlungsspezifische Kenndaten errechnet werden, die im ursprünglichen Sinne für Neubaugebiete festgelegt werden, lässt sich somit im Umkehrschluss die potentielle Einwohnerdichte aufgrund eines Bebauungsplans abschätzen.

Vorgegangen wird dabei nach dem folgenden Muster: in Kapitel 2 *Theoretische Grundlagen und Methoden* werden die klassischen Verfahren der Wasserbedarfsprognose mit aktuellen Forschungen verglichen. Der Stellenwert der angeschlossenen Verbraucher als Faktor auf den Wasserverbrauch wird untersucht. Es wird ein kurzer Überblick über Siedlungsstrukturen und verwendete Kennzahlen gegeben. Die Gesetzmäßigkeiten der Raumordnung werden abgehandelt. Das Kapitel wird durch die Beschreibung der Methoden zur Datensynthese und zur Datenanalyse abgeschlossen. In Kap. 3 *Datengrundlage* wird die verwendete Datengrundlage vorgestellt. In Kap. 4 *Datenaufbereitung und -analyse* wird das praktische Vorgehen zur Synthese und Analyse dokumentiert. Die Ergebnisse werden schließlich in Kap. 5 *Ergebnisse und Diskussion* diskutiert.

Die Synthese und Analyse wird in Kap. 4 *Datenaufbereitung und -analyse* beispielhaft für die Gemeinde Alpen im Kreis Wesel durchgeführt. Dies ist eine Kleinstadt mit ca. 13.000 Einwohnern. Zusätzlich werden die Berechnungen für die Stadt Solingen, die eine Einwohnerzahl von ca. 150.000 besitzt, sowie für die mittelgroße, nicht kreisfreie Stadt Lemgo mit ca. 41.000 Einwohnern durchgeführt. Diese Städte wurden aufgrund ihrer Einwohnerzahlen ausgewählt, so dass repräsentativ für jeweils eine große, eine mittelgroße und eine kleine Gemeinde die Berechnungen und Analysen durchgeführt werden.

2. Theoretische Grundlagen und Methoden

Vor der Bearbeitung des eigentlichen Themas wird eine Literaturrecherche durchgeführt. In Abschnitt 2.1 *Wasserbedarfsprognose* werden zunächst Informationen über die Erstellung von Wasserbedarfsprognosen und den aktuellen Forschungsstand diesbezüglich gesammelt. Das Hauptaugenmerk liegt dabei auf dem Einfluss der Verbraucherzahlen (Einwohnerdichten) auf die Erstellung einer integrierten Wasserbedarfsprognose. In Abschnitt 2.2 *Siedlungsstruktur* werden Grundlagen über die Themen Siedlungsstruktur und -formen, siedlungsstrukturelle Kennzahlen und die Raumordnung aus der Literatur gesammelt. Ebenfalls wird über die Eigenschaft der Bodenrichtwerte als sozio-ökonomischer Faktor berichtet. Dieser erste Teil des Kapitels stellt somit die theoretischen Grundlagen dar.

Im zweiten Teil dieses Kapitels werden die Methoden benannt. Diese sind die Aufgaben von Geoinformationssystemen und der statistischen Analyse. Dazu werden benötigte Begriffe erklärt. Es wird auf die jeweils verwendete Software eingegangen.

Dieses Dokument wurde erstellt mit dem Textsatzsystem \LaTeX . Als Editor dient `kile`. Die Literaturverwaltung wurde mit `biblatex` durchgeführt; als grafisches Frontend kommt `kbibtex` zur Anwendung.

2.1. Wasserbedarfsprognose

Traditionelle Wasserbedarfsprognosen setzen einen hohen Kenntnisstand über die angeschlossenen Verbraucher in einem zu erschließenden Gebiet voraus. Fritsch et al. (2014) ist der „Klassiker“ im Bereich Wasserversorgung. Das Werk wird von allen möglichen Zielgruppen herangezogen und ergänzt dabei das Regelwerk des DVGW und bezieht sich z.T. auch auf dieses. Für die Planung von Trinkwasserinfrastrukturen sind detailreiche Angaben über die Art und Zusammensetzung von Verbrauchern notwendig. Hierfür wird ein Tabellenwerk bereitgestellt, an dem sich der jeweilige Planer orientieren kann. Allerdings sind aufgrund von strukturellen und klimatischen Unterschieden Schwankungen von 15% und mehr durchaus möglich (ebd., S. 42). Das Tabellenwerk (ebd., S. 42 ff.) listet verschiedene Verbraucher auf, z. B. Haushalt einschließlich Kleingewerbe und Landwirtschaft, Industrie einschließlich Großgewerbe sowie sonstige Verbraucher wie öffentliche Einrichtungen.

Karger, Cord-Landwehr und Hoffmann (2008) ist ein im Vergleich dazu relativ kleines Werk. Es geht auf Verbrauchskennwerte (ebd., S. 11) und die kleinteilige Berechnung mit bekannten Verbraucherzahlen ein und verweist dabei auf die BGW-Wasserstatistik, die nicht mehr unter diesem Namen, sondern als BDEW-Statistik weitergeführt wird.

Hiessl (2001) gibt einen detaillierten Überblick über das Thema Wasserbedarf und über den Sinn und Nutzen von Wasserbedarfsprognosen. Der Autor listet zudem Faktoren auf, die den Wasserbedarf beeinflussen. Ein Teil dieser Faktoren wird auch von den aktuellen Studien (siehe Abschnitt 2.1.1) verwendet. Zudem beschreibt er den „Paradigmenwechsel“, der sich im Bereich Wasserverbrauch zur Zeit vollzieht.

Unter dem Titel „Wasserbedarf – Ein Konzept im Wandel“ weist er zunächst auf die Gesetze und Normen hin, die diesem Thema zu Grunde liegen. Er gibt eine sehr detaillierte Liste mit Faktoren an, die den Wasserbedarf kurz-, mittel- und langfristig beeinflussen. Zusätzlich beschreibt er noch den Paradigmenwechsel, der sich in Politik, Wissenschaft und Gesellschaft zum Thema Wasserbedarf und -verbrauch vollzieht.

Der Autor macht deutlich, dass Wasser in zahlreichen Nutzungsprozessen eine zentrale Rolle spielt. Dazu gehören u.a. die Wirtschaft sowie der öffentliche und private Bereich. In den wasserreichen Industrieländern sind Grundkonzepte für Infrastrukturen zur Ver- und Entsorgung von Wasser entstanden, die eine lange Tradition haben und sich nicht an den technologischen Fortschritt angepasst haben. Ebenso wenig angepasst haben sich die Wasserbedarfsprognosen, die die Grundlage für Planung und Betrieb dieser Infrastruktursysteme darstellen, was zur Folge hat, dass diese über- oder unterdimensioniert sind und nicht mehr den aktuellen Anforderungen gerecht werden. ebd., S. 44

Da aber das Wasserhaushaltsgesetz (WHG) vorschreibt, dass Gewässer nachhaltig zu bewirtschaften sind mit dem Ziel, „sie zum Wohl der Allgemeinheit und im Einklang mit ihm auch im Interesse

Einzelner zu nutzen”^{2.1}, sind alle verpflichtet, „eine mit Rücksicht auf den Wasserhaushalt gebotene sparsame Verwendung des Wassers sicherzustellen“^{2.2}. Dies ist nicht nur auf die Verbraucher zu beziehen; der Gesetzgeber schreibt damit auch den Wasserversorgern vor, auf Sparsamkeit zu achten. Ein überdimensioniertes System, dass von Zeit zu Zeit gespült werden muss, ist nicht effektiv und sparsam (vgl. Kluge, Libbe und Schramm 2006).

Die Deutsche Industrienorm (DIN) 4046 definiert den Begriff des Wasserbedarfs als einen Planungswert für die Wasserversorger, um Trinkwasserinfrastrukturen in einem bestimmten Versorgungsgebiet nach dem voraussichtlichem Bedarf innerhalb eines bestimmten Zeitrahmens zu konzipieren, wobei die lokalen Gegebenheiten miteinbezogen werden müssen. (Hiesl 2001, S. 45)

Für die realistische Abschätzung des Wasserbedarfs führt Hiesl (ebd.) eine sehr ausführliche Liste an Faktoren zu Verbrauchscharakteristiken und zukünftigen Entwicklungen an (ebd., S. 46):

- Zahl der zu versorgenden Einwohner und demografische Entwicklung
- Verbrauchsgewohnheiten und Lebensstandard der Bevölkerung (auch techn. Stand der sanitären Anlagen)
- Einwohnerdichte, Gästebettdichte
- Eigen- und Einzelwasserversorgung
- Klima und Wetter
- Erschließung, Bebauung und Größe von Grundstücken, Gartenflächen
- Ausbau der Kanalisation
- Wasserbedarf von öffentlichen Einrichtungen, Gewerbe- und Industriebetrieben
- Verbraucher mit intensivem Wasserbedarf (Papierindustrie, Gärtnereien)
- Landwirtschaft (Ackerbau, Viehwirtschaft)
- Preisgestaltung, Abwassergebühren
- Löschwasserbedarf

Die Entwicklung der einzelnen Faktoren kann dabei quantitativ und zeitlich sehr unterschiedlich ausfallen, so dass sich einzelne Faktoren gegenseitig verstärken oder ausgleichen können. Erschwert wird die Einschätzung des Wasserbedarf dadurch noch zusätzlich, dass Infrastrukturen für einen Zeitraum von 40 bis 80 Jahren geplant und gebaut werden und Fehleinschätzungen einen hohen nachträglichen Kostenaufwand bedeuten können. (ebd.)

^{2.1}§6 (1) WHG, vgl. Berendes, Frenz und Müggenborg 2017

^{2.2}§5 (1) WHG, vgl. ebd.

Der von Hiessl (2001) beschriebene Paradigmenwechsel besteht darin, dass der Wasserbedarf nicht mehr nur allein angebotsorientiert betrachtet wird, sondern zunehmend nachfrageorientiert ist. Auch wird das Wasser nicht mehr als einmalig nutzbar angesehen, sondern kann je nach Verschmutzungsgrad noch weiteren Nutzungsarten zugeführt werden (z.B. Grauwasseraufbereitung, Kreislaufwirtschaft). Regenwasser und Abwasser werden nicht aus dem Einzugsgebiet abgeführt, sondern einer Nutzung zugeführt (Regenwasseraufbereitung, Grundwasseranreicherung). Die Dezentralisierung der Infrastruktur bringt zudem ökonomische und ökologische Vorteile. Durch die gemeinsame Betrachtung von Trink- und Abwasserinfrastruktur können Synergieeffekte sichtbar werden (z. B. Wärmerückgewinnung aus Abwasser). Standardlösungen treten in den Hintergrund; problemspezifische Lösungen bringen mehr Effizienz und Flexibilität mit. Die Akteure freunden sich schneller mit innovativen Lösungen an, was unter anderem auch förderlich für den Wettbewerb ist. Zusätzlich spielt die Öffentlichkeit eine immer größere Rolle. (ebd., S. 47 f.)

Malz und Scheele (2011) beschreiben den Unterschied zwischen veralteten Methoden der Wasserbedarfsprognose wie das „fixed-coefficient model“ (ebd., S. 145) und neueren Methoden, die auf Szenarien basieren (ebd., S. 146).

Wasserbedarfsprognosen sind ein wichtiges Instrument zur nachhaltigen Planung zukünftiger Wasserinfrastrukturen. In der Vergangenheit wurden solche Prognosen trendbasiert durchgeführt. Historische Verbrauchsdaten, also Daten aus der Vergangenheit, wurden dazu analysiert und auf die Zukunft projiziert. Im Wesentlichen basieren diese Prognosen auf der Annahme, dass ein linearer Zusammenhang zwischen dem gesamten Wasserverbrauch in einem Gebiet und der Bevölkerungsanzahl in dem selbigen besteht. Es wird also von einem spezifischen Wasserbedarf pro Person ausgegangen, der sich mit Hilfe der Bevölkerungszahl auf den Gesamtverbrauch hochrechnen lässt. Dieses Modell wird als „fixed-coefficient model“ bezeichnet (Renzetti 2005).

Nun hat sich rückblickend gezeigt, dass diese Methode bei keiner Prognose ein der Realität entsprechendes Ergebnis geliefert hat. Der von der Bundesrepublik Deutschland in den 70er und 80er Jahren des vergangenen Jahrhundert vorausgesagte Pro-Kopf-und-Tag-Verbrauch von 219 L für das Jahr 2000 wurde deutlich von der Realität unterschritten, denn er lag dann tatsächlich nur bei 136 L pro Kopf und Tag. Als Grund dafür wird von Malz und Scheele (2011, S. 145) angeführt, dass sich „verändernde ökonomische, ökologische und technische Rahmenbedingungen nur schwer in die traditionellen Prognosen integrieren lassen“.

Die heutige Methode der Wahl ist die Verwendung von Szenarien. Hier werden verschiedene Rahmenbedingungen (Klima, Bevölkerungsentwicklung, technologischer Fortschritt, Wertewandel) angenommen und dann jeweils ein Szenario berechnet. Für die Entwicklung des globalen Wasserbedarfs ist vor allem die Bevölkerungsentwicklung die treibende Kraft. Der Klimawandel wird vor allem in der Landwirtschaft (Bewässerung) aber auch in der Industrie (Kühlung) den Wasserbedarf steigern. Der

Wasserbedarf von Privathaushalten wird sich nur wenig ändern. (Malz und Scheele 2011, S. 146)

Dazu schreiben Schleich und Hillenbrand (2007): „In general, the relatively small or insignificant parameter estimates associated with rainfall may be rationalized by the small share of gardening in total residential water consumption in Germany compared to countries like the US or Spain.“ Das bedeutet, dass die ohnehin schon wenigen klimasensiblen Nutzungsarten für das Trinkwasser, hier die Gartenbewässerung, in Deutschland aufgrund der Struktur, also der geringen Anteils von Gärten an den Siedlungsflächen, einen geringen Einfluss auf den gesamten Wasserverbrauch haben. (Malz und Scheele 2011, S. 146)

2.1.1. Aktuelle Forschungen

Die Überlegungen von Kluge, Libbe und Schramm (2006) zu den Auswirkungen des demographischen Wandels auf die Wasserversorgung beziehen sich vor allem auf einen Schrumpfungstrend. Als Beispiel wird hier ein Rückgang in Nutzung und Verbrauch aufgrund von Wanderungsbewegungen in Ostdeutschland in den letzten 25 Jahren angeführt. Die Wanderungsbewegungen können dabei kleinräumiger („Entdichtung von Versorgungsgebieten bei gleichzeitiger flächenmäßigen Zuwachs versorgter Gebiete durch Zersiedelungseffekte im Umland“ (ebd., S. 2)) und großräumiger („Abwanderung in andere Regionen und Entdichtung bisheriger Versorgungsgebiete“ (ebd.)) Natur sein. Des weiteren werden ein Bevölkerungsrückgang aufgrund niedrigerer Geburtenraten, eine niedrigere Nachfrage von Industrie und Gewerbe an die Versorgungsinfrastruktur sowie eine Netzverkleinerung aufgrund des Rückbaus von Gebäuden oder Siedlungsteilen genannt (ebd.). Die Herausforderung besteht nun darin, die Versorgungsinfrastruktur, die seit Beginn des Industriezeitalters auf einen Durchsatz von 150 L pro Einwohner und Tag ausgelegt wurde und der dabei die Möglichkeit der Steigerung implementiert wurde (ebd., S. 3), jetzt so anzupassen, dass sie mit den zurückgehenden Verbräuchen umgehen kann. Das gestaltet sich als schwierig, da im Laufe der Zeit Pfadabhängigkeiten entstanden sind, auch aufgrund der vornehmlich zentralen Infrastruktur, und die Entwicklung der Infrastruktur weder ökonomisch noch technisch einer allgemeinen Kritik unterzogen worden sind (ebd., S. 4)

Kluge, Libbe und Schramm (ebd.) weisen darauf hin, dass „Verbrauchsrückgang und demographische Entwicklung nicht als monokausal bzw. deterministisch verknüpft betrachtet werden können“. Dies liegt darin begründet, dass ein deutlich sinkender Wasserverbrauch im Gebiet der Wasserversorgung dazu führt, dass dieser Schrumpfungstrend noch zusätzlich verstärkt wird. Dabei wird der spezifische Wasserverbrauch, also der Verbrauch pro Person, auch noch durch die Verwendung wassersparender Technologie sowie durch eine Änderung im Bewusstsein und damit wassersparendem Verhalten der Verbraucher gesenkt (ebd., S. 2). Da kein deterministischer Zusammenhang hergestellt werden kann, stellt sich die Frage, ob denn ein empirischer Zusammenhang zwischen dem personenspezifischen

Verbrauch und den Auswirkungen neuerer wassersparender Technologie besteht. Hier fehlen bis dato allerdings ausreichende Studien und Grundlagen (Kluge, Libbe und Schramm 2006, S. 3).

Die zurückgehende Wassernutzung hat quantitative und in der Folge auch qualitative Aspekte. Zunächst muss festgehalten werden, dass die Infrastruktur der Wasserversorgung Funktionsschwellen besitzt, bei deren Unter- bzw. Überschreitung eine Anpassung vorgenommen werden muss. Die Betreiber der Infrastruktur, die auf Steigerung und Wachstum konzipiert wurde, stehen jetzt vor der Aufgabe, eine Unterschreitung zu verkraften, die durch Mindestfließmengen und Mindestfließgeschwindigkeiten festgelegt ist. So muss zusätzlich Wasser in das Netz eingespeist werden, um die Leitungen zu spülen, sonst tritt in den Versorgungsleitungen Stagnation mit allen seinen qualitativen Folgen auf (ebd., S. 5). Es liegt auf der Hand, dass damit der eigene Wasserbedarf eines Versorgers im Verhältnis zur Menge des abgegebenen Wassers zunimmt und damit auch die Betriebskosten steigen.

Vor allem dort, wo die Rohrdurchmesser der Versorgungsleitungen sehr groß dimensioniert worden sind, um beispielsweise auch die Versorgung mit Löschwasser über das Leitungsnetz zu ermöglichen, führt der Rückgang des Wasserverbrauchs zu problematisch niedrigen Durchflussgeschwindigkeiten, so dass sich Zonen der Stagnation und Sedimentation ausbilden können (ebd., S. 7). Chemische Reaktionen, die eigentlich zu lange dauern, können nun ablaufen, so dass es zu Korrosion und in Folge dessen zu Rost im Wasser kommen kann (ebd.). Angeblich begünstigt Stagnation eine Wiederverkeimung des Wassers nach Rohrbrüchen und Reparaturarbeiten, weshalb vorbeugende Rohrnetzspülungen und Desinfikationen des Trinkwassers durchgeführt werden (ebd.). Es hat sich allerdings gezeigt, dass, so lange Keime bzw. Bakterien in einem Biofilm fixiert sind, diese einen positiven Einfluss auf die Qualität des Trinkwassers haben, da sie verwertbare Inhaltsstoffe im Wasser abbauen, die dann nicht mehr für mobile Keime im Wasser zur Verfügung stehen, sie selbst aber nicht in das Wasser gelangen (ebd., S. 8). Durch Spülungen und Desinfizierungen können diese Biofilme allerdings zerstört werden und die in ihnen enthaltenen Bakterien werden freigesetzt. Ähnlich verhält es sich mit den Spülungen bei Sedimenten. Sedimente werden durch Spülungen aufwirbelt und die darin enthaltenen Bakterien werden im Leitungsnetz freigesetzt (ebd.). Stagnation hat auch dann eine schlechte Auswirkung auf die Qualität, wenn nach längeren Standzeiten (mehrere Wochen) und dem dadurch verringerten Nährstoffangebot für die Mikroorganismen dann wieder Wasser mit verwertbaren Inhaltsstoffen zu ihnen gelangt: es kommt zu einer Wiederverkeimung (vgl. Wricke 2005). Dies alles hat qualitative Folgen: eine Verkeimung des Trinkwassers mit einer einhergehenden Gefahr der Infektion und eine geschmackliche Verschlechterung.

Ökonomische, soziale und ökologische Folgewirkungen sind nicht zu vermeiden. Durch den höheren Aufwand, den ein Versorger betreiben muss, steigen die Betriebskosten durch den Einsatz von mehr Materie und Energie. Die höheren Fixkosten werden auf den Verbraucher umgelegt, der Grundpreis für Trinkwasser steigt. Dadurch verschlechtert sich auch die soziale Akzeptanz. Als Monopolist, der sich weder ausdehnen noch abwandern kann, hat ein Versorgungsunternehmen eine sehr geringe

Flexibilität bei der Planung. Einsparungen in der Kapazität sind zudem kaum möglich (Kluge, Libbe und Schramm 2006, S. 13). So bleibt nur die weitere Erhöhung der Umlage von Fixkosten auf den Verbraucher, was wiederum zu einer Abwanderung führt, und somit zu weiteren Problemen für das Versorgungsunternehmen (ebd., S. 14).

Im Bereich der Wasserversorgung bedeutet paradoxer Weise ein Rückgang des Ressourcenverbrauchs eine Verschlechterung im ökologischen Sinne. Da aufgrund des geringeren Trink- und somit auch Rohwasserbedarfs Brunnenanlagen stillgelegt werden, werden auch die sie umgebenden Wasserschutzgebiete aufgelassen. Der Versorger wird von seiner bisher wahrgenommenen Verantwortung befreit. Die Belastung des Grundwassers durch die Landwirtschaft steigt, da zum einen die Einschränkungen eines Wasserschutzgebietes nicht mehr gelten, zum anderen auch Ausgleichszahlungen von Versorgern an Landwirte wegen verringerten Ertrages für geringeren Einsatz von Düngemitteln wegfallen (ebd.).

Weitere ökologische Folgen können darin bestehen, dass durch Korrosionsprozesse und anaerobe Umsetzungsprozesse chemische Verbindungen (z. B. Schwefelsäure, Schwefelwasserstoff) freigesetzt werden, die in das Grundwasser oder an die Oberfläche gelangen und sich dort negativ auswirken. Der häufigere Einsatz von Mitteln zur Desinfektion (Chlor etc.) hat im allgemeinen schlechte Auswirkungen auf den gesamten Wasserkreislauf (ebd., S. 15).

Grossmann und Hofmann (2008) und Kluge, Deffner et al. (2008) beschreiben zusammen das Prognose-Modell für Hamburg Wasser. Grossmann und Hofmann (2008) sehen die Wasserbedarfsprognose als Grundlage für folgende Planungsprozesse: Investition und Aufwand (Ergebnisprognose), Planung und Dimensionierung des Trinkwasserversorgungssystems sowie des Abwasserentsorgungssystems, Optimierung von Unterhaltungskonzepten und die wasserwirtschaftliche Planung (Bewertung des Grundwasserangebots). Dabei gehen sie davon aus, dass der Wasserbedarf abhängig ist von der Entwicklung von Gewerbe und Dienstleistungen, dem demografischen Wandel sowie Veränderungen des Klimas und ergänzen damit bisherige Modelle, die laut Grossmann und Hofmann (ebd., S. 759) „lediglich den Einfluss technischer Entwicklungen und städtebaulicher Faktoren auf die Entwicklung des Wasserbedarfs“ berücksichtigen, um eine Klassifizierung der Haushalte, die bisher als homogen angesehen wurden, und weitere sozioökonomische Faktoren wie das Verhalten von Verbrauchergruppen. Durch eine flächendeckende Verfügbarkeit der Daten auf Grundstücksebene und die Installation von Wasserzählern in den Haushalten konnte eine genaue statistische Analyse angefertigt werden, die aufgrund ihrer Differenzierung nach Flächen und Verbrauchergruppen auf andere Versorgungsgebiete übertragbar ist. Als Faktoren, die den Wasserverbrauch beeinflussen, werden die folgenden genannt: Wohnungswasserzähler (Selbstkontrolle des Verbrauchers), Wohnungsmodernisierung und

(wassersparende) Sanitärausstattung, (sparsame) Haushaltsgeräte, Haushaltsgröße (Personen), demografische Entwicklung, spezifische Wohnfläche, Gartenfläche sowie das Sparverhalten der Verbraucher. Bis 2030 rechnet man hier mit einem Rückgang von 7,5 % des absoluten Haushaltswasserbedarfs gegenüber dem spezifischen Wasserbedarf aufgrund von Bevölkerungswachstum durch Zuwanderung; im Jahr sind das durchschnittlich 1 %, die sich aber auf 0,3 % pro Jahr abschwächen werden. Für Gewerbe und Dienstleistungen geht man von einer Steigerung von 8,2 % aus; die Werte für Industrie und andere stagnieren. Dies sind allesamt gemittelte Werte, deren Schwankung bei 4,6 % liegen kann, wenn der Einfluss der Witterung (trockene oder nasse Jahre) besonders stark ist. (Grossmann und Hofmann 2008, S. 759 ff.)

Zusammenfassend kommt die Studie zu dem Ergebnis, dass der mutmaßliche Klimawandel den durchschnittlichen Wasserbedarf nicht beeinflusst, sondern nur zu Schwankungen aufgrund von besonderer Witterung und damit zur Erhöhung von Spitzenfaktoren im Wasserverbrauch führt. Dies ist vor allem auf die intensive Bewässerung von Gärten und Grünflächen in Trockenperioden zurückzuführen. Die o.g. Faktoren dagegen beeinflussen den spezifischen Wasserbedarf wesentlich. Da die Prognose auf der Ebene von Stadtteilen und Versorgungsgebieten flächig differenziert erfolgte, können teilsräumliche Entwicklungen getrennt von der allgemeinen Entwicklung abgebildet werden. (ebd., S. 762 f.)

Die Methodik, die der Wasserbedarfsprognose nach Grossmann und Hofmann (ebd.) zugrunde liegt, wurde von Kluge, Deffner et al. (2008) konzipiert. Diese integrierte Wasserbedarfsprognose (WBP), die stark differenzierte und heterogene Daten vereint, zeichnet sich durch die Verwendung einer teilsräumlichen Analyse, der Berücksichtigung von technischen, siedlungs- und sozialstrukturellen Faktoren, der gesonderten Berechnung für Industrie, Gewerbe und Dienstleistungen und einem integrierten Prognosemodell aus, das mit Hilfe einer Tabellenkalkulation erstellt wurde. Das methodische Vorgehen besteht hier darin, Verbrauchergruppen stark zu differenzieren, genaue Daten zu Wasseraufkommen und -abgabe der Versorger zu sammeln, Expertengespräche zur langfristigen Entwicklung zu führen, siedlungsstrukturelle Aspekte miteinzubeziehen, die Verbraucher in privaten Haushalten und im Sektor Gewerbe/Dienstleistung/Industrie in repräsentativer Zahl zu Wasserverbrauch und -nutzung zu befragen und die gesammelten Daten dann auszuwerten und mit Annahmen und Hypothesen zu vergleichen. Die Differenzierung in Verbrauchergruppen auf der obersten Ebene wird in Haushalte, Industrie, Gewerbe, öffentliche Einrichtungen, sonstige Verbraucher sowie Eigenbedarf der Versorger und Verluste vorgenommen, wobei die einzelnen Klassen noch weiter unterteilt werden. (ebd., S. 766)

Für die Beschreibung der Struktur eines Versorgungsgebietes sind die folgenden Institutionen als Lieferanten für Daten genannt, die in Tabelle 2.1 aufgelistet sind.

Zusätzlich wurden aus der Analyse der Expertengespräche Daten und Trends zur Bevölkerungsent-

Institution	Datengrundlage
Kommunale Behörden	Flächennutzungspläne, Stadtentwicklung
Regionale Behörden	Reg. Entwicklungskonzepte, Anforderung an WBP
Statistische Ämter	Sozialstruktur, Bevölkerungsentwicklung
Statistisches Bundesamt	Wirtschaftsrechnungen, Übersichten
Industrie- und Handelskammern	adressbezogener, gewerblicher Wasserbedarf
Große Wohnungsunternehmen	Bestand, Entwicklung; wasserbezogene Sanierungsstände
Forschungsinstitute	Berichte und Analysen zu Entwicklungstrends, frühere Gutachten
Umweltbundesamt/Wetterdienste	Meteorologische Daten, Klimaprognosen

Tabelle 2.1: Institutionen als Lieferanten für Daten nach Kluge, Deffner et al. 2008, S. 767

wicklung, Haushaltsgrößen, Wohnraum, Anzahl der Haushalte und Arbeitsplätze sowie zur Wirtschaftsstruktur abgeleitet (Kluge, Deffner et al. 2008, S. 767 f.). Siedlungsstrukturelle Aspekte wie z. B. Bruttogeschossflächen, bebaute und versiegelte Flächen, Einwohnerzahlen etc. werden hier durch ein Geoinformationssystem (GIS) miteinander verschnitten, d. h. diese GIS-Daten besitzen Koordinaten und können dadurch in einen räumlichen Bezug gesetzt werden (ebd., S. 768). Für die direkte Erfassung des Wasserverbrauchs werden sozialempirische Erhebungen durchgeführt. Private Haushalte und gewerbliche Verbraucher werden hierfür befragt (ebd.). Für den privaten Wasserbedarf wurden technische, siedlungs- sowie sozialstrukturelle Faktoren identifiziert, für den gewerblichen dagegen die Anzahl der Beschäftigten und wirtschaftliche Kenndaten (ebd., S. 769). Die eigentliche Prognose erfolgt modular, d. h. es werden separate Teilprognosen erstellt, die beispielsweise unterteilt sind in den Wasserbedarf von städtischen und ländlichen Gebieten oder privater und gewerblicher Verbraucher. Diese werden integriert zu einer Gesamtprognose des Versorgungsgebietes, die durch das modularisierte Vorgehen eine Aufschlüsselung nach Verbrauchergruppen erlaubt (ebd., S. 769 f.). Dadurch dass das Prognosemodell parametrisiert wurde, können unterschiedliche Szenarien betrachtet werden. Kluge, Deffner et al. (ebd.) spielen mögliche Entwicklungstrends in den Bereichen „Bevölkerungsentwicklung und demografischer Wandel“ (Altersstruktur, Haushaltsgröße), „Klimawandel“ (Temperatur, Niederschlag), „Verbraucherverhalten“ (Bewusstsein, Wasser sparen, Preisstruktur), „Gartenbewässerung“ (s. Klimawandel), und „technische Innovation“ (Kreislaufführung, neuartige Sanitärsysteme) durch. Das Nutzungspotential dieser Wasserbedarfsprognose liegt vor allem in der Optimierung der betrieblichen Vorgänge bei stagnierendem oder rückläufigem Wasserbedarf (vgl. Kluge, Libbe und Schramm 2006). Denn ein Wasserversorger muss sich zum Ziel setzen, die Kosten für sich und die Verbraucher zu verringern oder mindestens zu stabilisieren, die Wirkung auf die Umwelt möglichst gering zu halten sowie die Qualität seiner Dienstleistungen zu verbessern (Kluge, Deffner et al. 2008, S. 772). Weil die Datengrundlage auch aus betriebswirtschaftlicher Sicht von Bedeutung ist, hat auch der Wasserversorger ein gesteigertes Interesse, dass die gesamte notwendige

Datengrundlage zugänglich gemacht wird. Auf der anderen Seite kann dann ein Wasserversorger auch auf die empirischen Ergebnisse einer WBP zurückgreifen, um z. B. einen besseren Kundenservice anbieten zu können (vgl. Kluge, Deffner et al. 2008, S. 772).

Koegst und Tränckner (2010) beschreiben zwei Ansätze zur Wasserbedarfsprognose: einen semi-deterministischen und einen rein statistischen, wobei letzterer für die vorliegende Arbeit interessant ist, da hier der Pro-Kopf-Bedarf durch sozioökonomische Faktoren beschrieben wird. Die Autoren stellen fest, dass die langfristige Planung von Wasserversorgungsanlagen und Infrastrukturen bislang vor allem auf ein stetiges Wachstum der Bevölkerung setzte (ebd., S. 25). Dieser Trend ist nach neueren Erkenntnissen aber umgekehrt, so dass von einem Bevölkerungsrückgang mit gleichzeitiger Überalterung der Gesellschaft ausgegangen wird (ebd., S. 26). Allgemein werden als Gründe dafür der Rückgang in der Geburtenrate, die höhere Alterserwartung sowie Migration aus dem Ausland gesehen. Der Wegzug aus ländlichen Gebieten in die Städte und umliegenden Kreise verursacht weitere räumliche Unterschiede in der Einwohnerdichte, vgl. ebd. Als ersten Beleg für diesen Trend zitieren die Autoren die Raumordnungsprognose des Bundesamtes für Bauwesen und Raumordnung, welcher von einem leichten Anstieg der Bevölkerungszahlen in den alten Bundesländern ausgeht, einem deutlichen Rückgang in den neuen Bundesländern und damit einem insgesamt zurückgehenden Trend für die gesamte Bundesrepublik Deutschland (BRD) bis zum Jahr 2020 ausgehend vom Jahr 2002. Abgeschwächt wird dieser Trend durch Migration von außen (ebd., S. 26 f.). Die Überalterung betrifft die gesamte BRD, die alten Bundesländer allerdings stärker als die neuen. Damit ist der allgemeine Trend der Bevölkerungsdichte abgegrenzt, der aber nicht direkt auf den gesamten Wasserbedarf Rückschlüsse zulässt, da der Pro-Kopf-Wasserbedarf ebenso einer Veränderung unterworfen ist. Hierfür nennen Koegst und Tränckner (ebd.) Lebensgewohnheiten sowie wirtschaftliche und sozioökonomische Gründe wie ein gesteigertes Umweltbewusstsein und den Einfluss der Wasserpreise auf das Verhalten des Verbrauchers (ebd., S. 30 f.). So wie Kluge, Deffner et al. (2008) sehen sie aber auch die technischen Innovationen als eine treibende Kraft beim Wassersparen: die vermehrte Verwendung von wassersparenden Toiletten (NASS), Spül- und Waschmaschinen, sowie die Umstellung zu geschlossenen Wasserkreisläufen und die Substitution von Trinkwasser. Der Rückgang des Wasserverbrauchs ist demnach nicht nur der demografischen Entwicklung geschuldet, sondern auch dem Einsatz wassersparender Technologien.

Koegst und Tränckner (2010) machen darauf aufmerksam, dass der tatsächliche Pro-Kopf-Verbrauch sich eben nicht in dem vom DVGW^{2,3} durchschnittlichen Bereich von $120 \text{ L}/(E \cdot d)$ bewegt, sondern lokal stark variiert, was die Notwendigkeit eines konkreten Ansatzes zum Pro-Kopf-Verbrauch unterstreicht. Mit einer vorhersagbaren Bevölkerungsentwicklung und einem auf lokalen sozioökonomischen Faktoren beruhenden Pro-Kopf-Verbrauch kann daher eine Wasserbedarfsprognose für ein konkretes Versorgungsgebiet getroffen werden (ebd., S. 31). Es werden bei der Erstellung der Wasser-

^{2,3}Deutscher Verein des Gas- und Wasserfaches e. V. 2008

bedarfsprognose zwei Ansätze verfolgt: einen semi-deterministischen und einen rein statistischen, wobei letzterer für die vorliegende Arbeit interessant ist, da hier der Pro-Kopf-Bedarf durch sozioökonomische Faktoren beschrieben wird (Koegst und Tränckner 2010, S. 32). Die Autoren benutzen für diesen Ansatz Daten der Gemeinden aus Sachsen aus dem Jahr 2008. Diese sind allgemeine Merkmale zu Flächen und deren Nutzung, Einwohnerzahlen, Wohn- und Haushaltsgrößen, Gewerbe und Anzahl der Angestellten, Anzahl der Gästebetten im Tourismusgeschäft, öffentliche Einrichtungen sowie volkswirtschaftliche Merkmale. Zur Vergleichbarkeit werden diese Daten auf die jeweiligen Flächen bezogen. Anstatt im Voraus Klassen festzulegen, denen die Daten zugeordnet werden, wird mittels einer Clusteranalyse die Klassifikation anhand der Daten vorgenommen. In diesem Fall sind es vier Cluster, denen die Merkmale Einwohnerdichte, Gästebetten pro 100 Einwohner und Gewerbesteuer pro Einwohner zugeordnet werden. Nach einer Bereinigung der Daten (Wegfallen der Daten in Kreisen mit Verbräuchen von über 120 und mehr als 300 Gästebetten pro 1000 Einwohner) sowie durch das Fehlen von Daten in einzelnen Kreisen, wurden auf der Datenbasis von 204 Datensätzen (vorher 493) die folgenden Ergebnisse produziert: der mittlere Wasserbedarf liegt bei $85 L/(E \cdot d)$, wobei 50% der Gemeinden einen Wasserbedarf zwischen $68 L/(E \cdot d)$ und $86 L/(E \cdot d)$ haben (Median: $75 L/(E \cdot d)$) (ebd., S. 32 f.).

Insgesamt zeigte sich bei der Untersuchung von Koegst und Tränckner (ebd.), dass der Wasserverbrauch in kleinen Gemeinden sehr stark vom Tourismus geprägt ist, was bei größeren Städte keine Rolle mehr zu spielen scheint. Der Einsatz von sparsamen Geräten führt zu einer Absenkung des Wasserbedarfs, was aber durch die Haushaltsgröße (Single, Mehrpersonen) überlagert wird. Ebenfalls führten die Autoren eine Untersuchung zur Auswirkung des Klimas (Temperatur, Trockenheit) auf den Wasserverbrauch durch, wobei hier kein nennenswerter Anstieg, auch nicht bei den Spitzenverbräuchen, zu verzeichnen war, was allerdings für ländliche geprägte Versorgungsgebiete noch untersucht werden muss (ebd., S. 54 f.).

Schumann und Korth (2014) beschreiben den Zwischenstand ihrer Studien im Forschungsprojekt „Verbrauchsganglinien“ des Deutschen Verein des Gas- und Wasserfaches e. V. (DVGW). Im Rahmen des DVGW-Forschungsprojektes „Aktualisierung der Verbrauchsganglinien für Haushalte, öffentliche Gebäude und Kleingewerbe sowie Entwicklung eines Modells zur Simulation des Wasserbedarfs“ sind 13 Forschungsunternehmen und die TU Dresden involviert, um neue Prognosen bezüglich des Trinkwasserbedarfs in der Bundesrepublik Deutschland zu erarbeiten (ebd., S. 51). Da die in den 1970er Jahren prognostizierten Werte erheblich vom aktuellen Verbrauch abweichen, ist dies ein notwendiger Schritt, da bisher die Planungen für Trinkwasserinfrastrukturen auf Kennwerten basieren, die aus diesen nun als verfehlt zu betrachtenden Prognosen abgeleitet wurden (vgl. ebd., S. 53). Die Regelwerke, die seit den 1980er Jahren für die Auslegung von Trinkwasserinfrastrukturen zu Rate gezogen werden, sind nach Schumann und Korth (ebd.):

- Deutsches Institut für Normung: Technische Regeln für Trinkwasser-Installationen – Teil 300:

Ermittlung des Rohrdurchmessers (2012)

- DVGW: Arbeitsblatt W 400-1: Technische Regeln Wasserverteilungsanlagen (TRWV), Teil 1: Planung (2004)
- DVGW: Arbeitsblatt W 410: Wasserbedarf – Kennwerte und Einflussgrößen (2008)

Beispielsweise weist der nach diesen Regelwerken bemessene Spitzenbedarf von Wohngebäuden mit 100 Wohneinheiten eine Differenz von bis zu $11 \text{ m}^3/\text{h}$ auf (Schumann und Korth 2014, S. 53). Die durch die viel zu hoch angenommenen Werte für den Pro-Kopf-Bedarf führen im schlechtesten Fall zu einer Überdimensionierung mit Folgen von zu geringen Fließgeschwindigkeiten und zu langen Aufenthaltszeiten des Trinkwassers im Rohrnetz (ebd.).

Im Rahmen ihres Messprogramms führen Schumann und Korth (ebd.) insgesamt mehr als 250 Einzelmessungen in unterschiedlichen Verbrauchsstellen im gesamten Bundesgebiet durch (ebd., S. 54). Die Auswahl der Messstellen beruht auch auf der Betrachtung von sozioökonomischen, technologischen, infrastrukturellen und klimatischen Aspekten, ähnlich wie bei Kluge, Deffner et al. (2008). Dabei wurden die statistisch repräsentativen Gebiete aufgrund der Faktoren „aktueller Mietspiegel“, „Arbeitslosenquote“, „Altersquotient“, „Durchschnittliche Haushaltsgröße“ und „Bevölkerungsanzahl im Postleitzahlgebiet“ ausgewählt (Schumann und Korth 2014, S. 55).

An den Messstellen wurden die vorhandenen Hauswasserzähler durch Ultraschallzähler ersetzt, die eine sehr detaillierte Aufzeichnung des Durchflusses im Sekundenbereich ermöglichen. Durch die separate Aufzeichnung von Einzelverbrauchern konnten auch bei Messstellen mit mehreren angeschlossenen Verbrauchern Ereignisse identifiziert werden, wie zum Beispiel das Betätigen der Toilettenspülung oder der Wasserverbrauch während eines Waschprogramms einer Waschmaschine. Durch statistische Auswertungen wurden dann ordinäre Ganglinien, beispielsweise die eines Einfamilienhauses, konstruiert. Diese Ganglinien sollen in Zukunft Einfluss haben auf die Überarbeitung der angesprochenen Regelwerke. (ebd., S. 56 ff.)

In ihrem Abschlussbericht (Martin, Korth und Schubert 2017), der Ende 2017 veröffentlicht wurde, fassen die Autoren das Ergebnis ihrer Studie zum Wasserverbrauch in Wohngebäuden zusammen (ebd., S. 97): es gibt eine hohe Korrelation zwischen dem maximalen Tages- und Spitzenverbrauch und der Zahl der Wohneinheiten und ihrer Bewohner. Allerdings scheint dieser Zusammenhang nicht linear zu sein. Die Bewohnerdichte bzw. Haushaltsstruktur hat signifikanten Einfluss auf die Ausprägung der Spitzendurchflusswerte. Diese sind unterschiedlich ausgeprägt, sowohl bezogen auf die Region als auch auf die Jahreszeit. Dabei hat der Trend hin zu Ein-Personen-Haushalten eine merklich Auswirkung auf die Spitzenwerte. Des weiteren wurde die Beobachtung in dünn besiedelten Gebieten gemacht, dass eine gesteigerte Außentemperatur vereinzelt zu einem höheren Spitzendurchfluss führen kann. Das Vorhandensein von Kleingewerbe hat keinen deutlichen Einfluss auf den Spitzendurchflusswert.

Als Empfehlung geben die Autoren an, die Regelwerke des DVGW W 406 und W 410 aufgrund der neuen Ergebnissen zu prüfen und ggf. zu überarbeiten. Dabei soll darauf geachtet werden, „wie hoch der Anpassungsbedarf in den einzelnen Regelwerken ist, welche weiteren Regelwerke durch die Veränderung der Spitzendurchflusswerte betroffen sind, ob die Einbeziehung der Untersuchungsergebnisse hinsichtlich der Abhängigkeit von Anzahl der Einwohner und Bewohnerdichte zweckmäßig sind und welche Sicherheitsfaktoren zu berücksichtigen sind“ (Martin, Korth und Schubert 2017, S. 101). Aufgrund der Unterschiede in den Regelwerken und den Untersuchungsergebnissen geben die Autoren an, dass sie es als sinnvoll betrachten, weitere Untersuchungen vorzunehmen, um auch für nicht-häusliche Verbraucher bessere Empfehlungen abgeben zu können (ebd.).

Neunteufel, Richard, Perfler et al. (2012) haben eine umfangreiche Studie zu Verbrauchsganglinien in Österreich durchgeführt. Die Autoren kommen zu dem Ergebnis, dass zwar sozioökonomische Faktoren den durchschnittlichen Jahresverbrauch erheblich beeinflussen. Die Unterschiede im Wasserverbrauch liegen vor allem im täglichen Verbrauch und sind stark abhängig von Tageszeit und Wochentag und der Wetterlage. Das hat zur Folge, dass „die Variabilität des Verbrauchs aufgrund der kurzfristigen Einflussfaktoren [...] jene von langfristig wirkenden Faktoren jedenfalls deutlich [übersteigt]“ (ebd., S. 11). Die strukturellen Unterschiede dagegen wirken sich in Österreich viel stärker auf den Verbrauch aus, als dies durch klimatische Einflüsse bedingt würde (ebd.). Die Unterschiede im Wasserverbrauch zwischen ländlichen und städtischen Versorgungsgebieten liegt generell im jeweiligen Anteil des Gewerbes und der mitversorgten Industrie sowie öffentlichen Einrichtungen begründet. Bezogen auf die Wochentage ist aufgrund der Pendleraktivität von Arbeitnehmern an Werktagen in den Städten ein höherer Wasserverbrauch vorhanden als am Wochenende. In ländlichen Versorgungsgebieten verhält es sich genau umgekehrt (ebd.).

Selbstverständlich gibt es noch weitere Forschungsarbeiten, auf die nicht weiter eingegangen wird. Zumindest einige seien hiermit kurz erwähnt. Herber, Wagner und Roth (2008) benutzen für ihren Wasserbedarfsnachweis das konventionelle System der einfachen Pro-Kopf-Berechnung. Schleich und Hillenbrand (2007) vergleicht sein Modell mit anderen Regressionsmodellen zur Prognose des Wasserbedarfs. Desweiteren gibt es noch die Studien von Haffner (2004), Masannek (1996) und Winje und Iglhaut (1983).

Landgemeinde	2.000 – 5.000 E
Kleinstadt	5.000 – 20.000 E
Mittelstadt	20.000 – 50.000 E
Großstadt	über 100.000

Tabelle 2.2: Klassifizierung durch Einwohnerzahl nach Korda (2005, S. 81)

2.2. Siedlungsstruktur

Die Strukturen von Siedlungsgebieten sind über Jahrhunderte in der Gesellschaft entstandene Formen. Dazu gibt es noch Ordnungen und Gesetze, die die Veränderung und den Auf- und Abbau von Siedlungen lenken. Im Folgenden werden Eigenschaften von Siedlungsgebieten beschrieben, die historisch entstanden sind und die von Gesetzen vorgeschrieben sind. Über den Bodenrichtwert ist zusätzlich eine ökonomische Einflussnahme in die Besiedlung von Gebieten gegeben.

2.2.1. Siedlungsformen, Klassifizierung

Korda (2005) und Albers und Wékel (2011) geben einen Überblick über das Thema Städtebau. Albers und Wékel (ebd.) machen darauf aufmerksam, dass Stadtplanung immer eine praxisorientierte Disziplin darstellt. Dabei ist Erfahrung eine wichtige Voraussetzung für den Erfolg, da die Wirkungen von Stadtplanung im Vorfeld richtig abgeschätzt werden müssen (ebd., S. 17). Daher entspricht die Theorie der Stadtplanung niemals den Ansprüchen an naturwissenschaftliche Theorien, da keine Wiederholbarkeit von Experimenten als Untermauerung einer Theorie besteht (ebd.). Sie besteht vielmehr darin, Erfahrungen und Beobachtungen systematisch aufzubereiten. Nichtsdestotrotz muss im Städtebau die Realität durch Modelle dargestellt werden, deren Erfolg maßgebend davon abhängt, ob die richtigen Einflüsse für das Modell ausgewählt worden sind (ebd.).

Korda (2005) behandelt das Thema Städtebau. Dabei ist es ihm wichtig, bei allen Betroffenen der Planung, also die Planer und „Beplanten“, ein Verständnis dafür zu wecken, dass die gegenseitige Rücksichtnahme und das Wissen um die entstandenen und sich entwickelten Bedingungen für einen Interessenausgleich unabdingbar sind. Der Autor geht auf die wesentlichen Aspekte der städtebaulichen Planung und Gestaltung ein und behandelt auch die technischen Voraussetzungen für eine integrierte Planung.

Für die Raumordnung und Entwicklungsplanung ist es wichtig, Gemeinden strukturell zu untergliedern. Unterscheidungsmerkmale für die soziale und wirtschaftliche Struktur können dabei sehr vielfältig sein. Die einfachste Klassifizierung ist die Einwohnerzahl (E). Diese ist in Tabelle 2.2 aufgelistet. Hier fällt auf, dass der Autor für die Einwohnerzahl von 50.000 bis 100.000 keinen Typus

nennt. Wenn eine komplexere Unterteilung vorgenommen wird, die auch Aspekte aufnimmt, die im Bewusstsein der Einwohner vorhanden sind, spricht man dann umfassend von einem geografischen Stadtbegriff. Dieser berücksichtigt auch die Tatsache, dass durch Zusammenlegung von Verwaltungseinheiten zwar hohe Einwohnerzahlen, aber nicht der einer Stadt entsprechende visuelle Charakter vorliegt. (Korda 2005, S. 81)

Es gibt eine Reihe von Begriffen abseits der Typologie im Städtebau, die gesellschaftlich gewachsen sind. Dazu gehören Einzel- und Streusiedlung, Weiler, Dorf, Stadt, Großstadt (ebd., S. 81 f.). In Verdichtungsräumen, dort wo die kommunalen Gebietsgrenzen nicht mit den erkennbaren Siedlungsgebieten übereinstimmen, wird folgende Einteilung vom Kern zu den Randlagen hin vorgenommen: Stadtkern, Ergänzungsgebiet, verstärkte Zone, Randzone (ebd., S. 82).

Die übliche Maßeinheit der Fläche in Kartenwerken ist der Hektar (ha). Weniger häufig werden der Quadratmeter (m²) sowie der Quadratkilometer (km²) verwendet. Diese Maßeinheiten sind Grundlage für amtliche Kartenwerke. Diese sind unterteilt in Katasterpläne und Kataster-Plankarten (mit Maßstäben von 1:500 bis 1:5.000), topografische Karten (mit Maßstäben von 1:5.000 bis 1:20.000) sowie geografische Karten (mit Maßstäben bis 1:200.000). Daneben gibt es auch noch die thematischen Karten, die Informationen zu einem bestimmten Themenbereich „geo-grafisch“ darstellen. (ebd., S. 88) Als Gebietseinheiten gelten strukturell homogene und ausgeglichene Gebiete, die zum Zweck der Raumforschung in der Bundesrepublik Deutschland definiert worden sind. Es sind 97 an der Zahl. Beispielhaft genannt werden an dieser Stelle Bochum, Essen, Siegen, Paderborn, Bonn (ebd., S. 90).

Der Autor entwirft eine Gebäudetypologie, die eine Unterteilung von Gebäudeformen aufgrund der darin befindlichen Wohnformen vornimmt. Die Gebäudeformen sind die folgenden (ebd., S. 102 ff.): Die kleinste Form des Einzelhauses ist dabei das freistehende Einfamilienhaus, das sich durch einen großen Flächenverbrauch aufgrund des vorgeschriebenen freibleibenden Raumes zur Grundstücksgrenze hin auszeichnet. Es verbraucht eine Fläche von ca. 600 m². Die erzielbare Wohndichte liegt bei 30 bis 75 E/ha. Werden zwei Einfamilienhäuser „aneinander gebaut“ und somit auf die freibleibende Fläche zur Grundstücksgrenze verzichtet, spricht man von einem Doppelhaus. Dessen Werte liegen zwischen dem Einfamilienhaus und dem Reihnhaus. Das Reihnhaus ist quasi die Aneinanderreihung von Einfamilienhäusern. Pro Wohneinheit (We) ist hier mit einem Flächenverbrauch von 170 bis 220 m² zu rechnen. Die erzielbare Wohndichte liegt bei 150 bis 200 E/ha. Eine Sonderform des Reihenhauses ist das Kettenhaus. Hier liegen die Werte für die Grundstücksfläche bei 300 bis 500 m²/We und die erzielbare Nettowohndichte bei 80 bis 150 E/ha. Die größte Form des Einzelhauses ist dann schließlich das Mehrfamilienhaus, in dem Einwohnerdichten von 100 bis 120 E/ha üblich sind. Im Städtebau finden dann noch verschiedene Formen des Geschosswohnungsbaus Verwendung. Zu nennen sind hier das Außenganghaus mit 250 bis 400 E/ha, das Innenganghaus mit ca. 300 E/ha. Dazu gibt es noch eine Reihe von Sonderformen, zu denen keine Werte genannt werden.

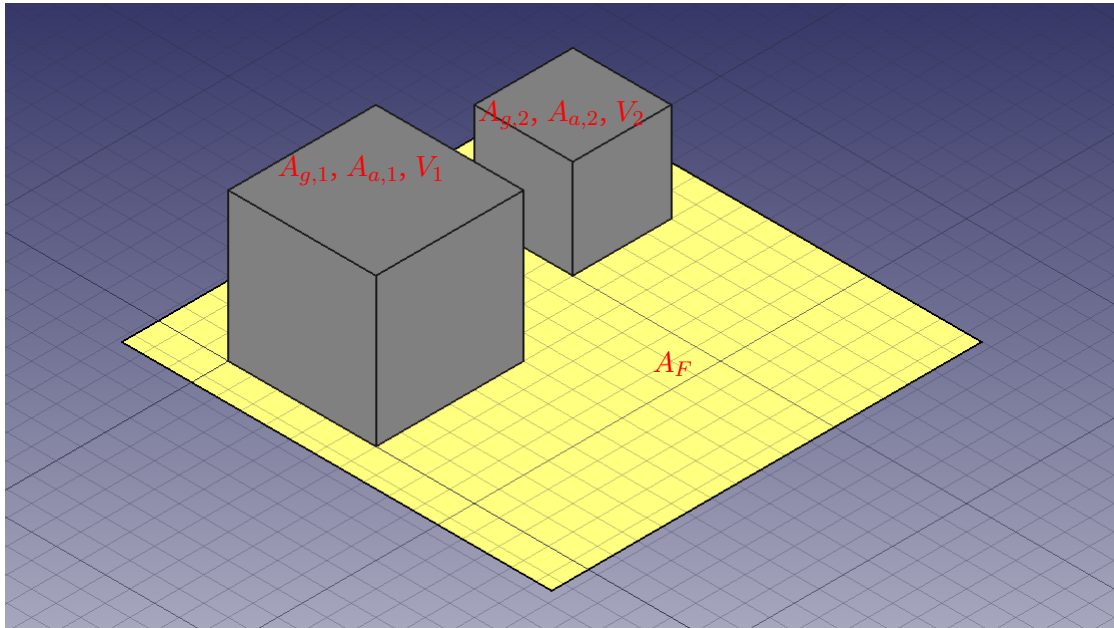


Abbildung 2.1: Beispielhafte Darstellung eines Flurstücks mit Gebäuden und Geometriedaten

Desweiteren werden von Korda (2005) noch die Begriffsdefinition nach der Baunutzungsverordnung (BauNVO) geklärt, wie z. B. Bruttobaugebiete, Nettobauland und noch weiter Dichtebegriffe und Dichtewerte wie die Bevölkerungsdichte, Wohndichte, Wohnungsdichte und Belegungsziffern aufgelistet. Auf diese wird an einer späteren Stelle in dieser Arbeit eingegangen. (ebd., S. 111 ff.)

2.2.2. Siedlungsstrukturelle Kennzahlen

Die verwendeten siedlungsstrukturellen Kennzahlen sind zum Teil in der DIN 277 und in der Baunutzungsverordnung (BauNVO) vorgeschrieben, siehe nächste Abschnitte. Zusätzlich wird noch eine weitere Kennzahl genutzt, die Kompaktheit. Alle Kennzahlen können aus den Geometriedaten der Grundstücke und Gebäude berechnet werden. Die Grundflächenzahl (GRZ), die Geschossflächenzahl (GFZ) und die Baumassenzahl (BMZ) sind relative Maße. Sie sind in ihrem Bezug zur Fläche des Baugrundstücks definiert (Boeddinghaus 2000, S. 229). In Abb. 2.1 ist beispielhaft ein Flurstück mit zwei Gebäuden dargestellt (eigene Darstellung). Das Flurstück ist hier eine einfache rechteckige Fläche. Die zwei Gebäude sind als Klötzchen dargestellt, so wie es für die gewählte Datengrundlage typisch ist, siehe Abschnitt 3.2 *3D-Gebäudedaten*.

Aus den Geometriedaten der Gebäude, wie der Außenfläche A_a , der Grundfläche A_g und dem Volumen V , der Höhe h und der Fläche des Flurstücks A_F lassen sich die folgenden siedlungsstrukturellen Kennzahlen berechnen.

Grundflächenzahl (GRZ) Die GRZ gibt den Flächenanteil eines Baugrundstückes an, der überbaut werden darf (Boeddinghaus 2000, S. 268). Die überbaute Fläche entspricht dann auch der Fläche der Dächer. Dazu heißt es in §19 BauNVO (1): „Die Grundflächenzahl gibt an, wieviel Quadratmeter Grundfläche je Quadratmeter Grundstücksfläche im Sinne des Absatzes 3 zulässig sind“ und weiter unter (2) „Zulässige Grundfläche ist der nach Absatz 1 errechnete Anteil des Baugrundstücks, der von baulichen Anlagen überdeckt werden darf“. In Abschnitt 3 heißt es weiter: „Für die Ermittlung der zulässigen Grundfläche ist die Fläche des Baugrundstücks maßgebend, die im Bauland und hinter der im Bebauungsplan festgesetzten Straßenbegrenzungslinie liegt. Ist eine Straßenbegrenzungslinie nicht festgesetzt, so ist die Fläche des Baugrundstücks maßgebend, die hinter der tatsächlichen Straßengrenze liegt oder die im Bebauungsplan als maßgebend für die Ermittlung der zulässigen Grundfläche festgesetzt ist.“ (ebd., S. 268 f.)

Die GRZ wird somit aus dem Verhältnis von der Summe aller Grundflächen A_g der Gebäude $0 \dots n$ zur gesamten Fläche des Flurstücks A_F berechnet. Die GRZ ist eine dimensionslose Zahl und wird daher als Dezimalbruch, selten in Prozent angegeben. Folgende Formel wird dabei angewendet.

$$GRZ = \frac{\sum_i^n A_{g,i}}{A_F} \quad (2.1)$$

Geschossflächenzahl (GFZ) Die GFZ gibt den Anteil der Flächen aller Geschosse der Gebäude an der Fläche eines Grund- bzw. Flurstücks an. Dabei gelten die folgenden Kriterien nach §19 BauNVO. Abschnitt 1: „Als Vollgeschosse gelten Geschosse, die nach landesrechtlichen Vorschriften Vollgeschosse sind oder auf ihre Zahl angerechnet werden.“ Die Definition der GRZ folgt in Abschnitt 2: „Die Geschossflächenzahl gibt an, wieviel Quadratmeter Geschossfläche je Quadratmeter Grundstücksfläche im Sinne des §19 Abs. 3 zulässig sind.“ Die Ermittlung erfolgt nach den Vorschriften in Abschnitt 3: „Die Geschossfläche ist nach den Außenmaßen der Gebäude in allen Vollgeschossen zu ermitteln. Im Bebauungsplan kann festgesetzt werden, dass die Flächen von Aufenthaltsräumen in anderen Geschossen einschließlich der zu ihnen gehörenden Treppenräume und einschließlich ihrer Umfassungswände ganz oder teilweise mitzurechnen oder ausnahmsweise nicht mitzurechnen sind.“ (ebd., S. 278)

Die GFZ wird somit aus dem Verhältnis von der Summe aller Geschossflächen der Gebäude $0 \dots n$ zur gesamten Fläche des Flurstücks A_F berechnet. Die Geschossflächen eines Gebäudes können unter der Annahme, dass Geschosse eine durchschnittliche Höhe von 3 m besitzen, aus der Höhe des Gebäudes h abgeleitet werden. Die GFZ ist ebenfalls eine dimensionslose Zahl, die in Prozent angegeben werden kann. Folgende Formel wird dabei angewendet.

$$GFZ = \frac{\sum_i^n A_{g,i} \cdot \frac{h_i}{3m}}{A_F} \quad (2.2)$$

Baumassenzahl (BMZ) Durch die BMZ wird ausgedrückt, wieviel Volumen bzw. Baumasse im Verhältnis zur Fläche eines Grundstück besteht. Die Definition ist ebenfalls in der BauNVO hinterlegt. §19 BauNVO (1): „Die Baumassenzahl gibt an, wieviel Kubikmeter Baumasse je Quadratmeter Grundstücksfläche im Sinne des § 19 Abs. 3 zulässig sind“ (Boeddinghaus 2000, S. 293). Wie die GRZ und die GFZ ist auch die BMZ ein auf die Grundstücksgröße bezogener relativer Maßfaktor (ebd.). Sie ist im Gegensatz zur GRZ und GFZ nicht dimensionslos, allerdings wird die Einheit m^3/m^2 bzw. m meist nicht aufgeführt. Sie berechnet sich zu:

$$BMZ = \frac{\sum_i^n V_i}{A_F} \quad (2.3)$$

Kompaktheit Als Maß für das Verhältnis zwischen umbautem Volumen und Außenflächen dient die sogenannte Kompaktheit. Sie ist das Verhältnis zwischen Fläche A und Volumen V eines Volumenkörpers. Eine andere Bezeichnung dafür ist der Formfaktor. Bezogen auf geometrische Körper hat eine Kugel das geringste Oberflächen-Volumen-Verhältnis, gefolgt von einem Würfel. Generell sinkt das Verhältnis bei Zunahme des Durchmessers bzw. der Kantenlänge. Bezogen auf architektonische Bauten besitzen frei stehende Einfamilienhäuser ein A/V -Verhältnis von 0,7 bis über 1,0, Doppelhäuser 0,6 bis 0,9, Reihenhäuser 0,4 bis 0,6 und Mehrfamilienhäuser einen Formfaktor von 0,3 bis 0,5 (Merbecks et al. 2015). Die relativ kleinen Werte gegenüber den Werten der geometrischen Körper lassen sich so erklären, dass nicht alle Außenflächen (Bodenfläche gilt nicht als Außenfläche) mit in die Berechnung dieser Formfaktoren eingehen. Die genannten Werte sind typische Werte. Sie hängen von der Größe des Gebäudes sowie von der Ausführung und Ausformung ab (ebd.). Die maximal mögliche Hüllfläche wird mit dem Faktor 0,9, die minimal mögliche Hüllfläche mit dem Faktor 0,3 errechnet.

Die Berechnung wird für alle Gebäude auf einem Grundstück durchgeführt, wobei die Summe aller Oberflächen, abzüglich der jeweiligen Grundflächen, im Verhältnis zur Summe der Volumina der Gebäude gesetzt werden, und erfolgt nach der Formel:

$$K = \frac{\sum_i^n A_{a,i}}{\sum_i^n V_i} \quad (2.4)$$

DIN 277 Ruf (2016) kommentiert die DIN 277 ausführlich und gibt einen gesamtheitlichen Überblick über siedlungsstrukturelle Kennzahlen. Die Neufassung der DIN 277-1 „Grundflächen und Rauminhalte im Bauwesen- Teil 1: Hochbau“ ist im Januar 2016 erschienen. Die bisherige Fassung war datiert auf den Februar 2005. Damit liegt eine sehr aktuelle neue Fassung vor. Die zeitliche Entwicklung der Norm geht bis auf das Jahr 1934 zurück. Damals hatte sie die Bezeichnung „DIN 277-1934-08 Umbauter Raum von Hochbauten“ (ebd., S. 451).

Zusammen mit der Norm DIN 276 stellt die DIN 277 die wichtigste technische Regel im Bereich der Bauökonomie dar und ist somit zentrales Instrument in den Bereichen Kommunikation, Information und Zusammenarbeit aller Beteiligten. Sie ist über die gesamte Lebensdauer eines Bauwerkes vom Entwurf bis hin zum Abbruch gültig und somit auch geeignet, Aussagen über die Einordnung und Klassifizierung von Gebäuden zu „untermauern“ (Ruf 2016, S. 4 f.).

Als „DIN 277“ wird allgemein die Norm „Grundflächen und Rauminhalte im Bauwesen“ bezeichnet. Hier ist dabei nicht festgelegt, aus welchem Jahr diese Norm stammt, sondern vielmehr ist damit die Norm als sich im Laufe der Zeit wandelndes und den aktuellen Herausforderungen angepasstes Regelwerk zu sehen. Die aktuell gültige Ausgabe ist die „DIN 277-1“ aus dem Januar 2016, die deshalb auch ausführlich mit „DIN 277-1: 01-2016“ bezeichnet werden kann. Im Folgenden wird aber nur die Kurschreibweise DIN 277-1 verwendet (ebd., S. 448).

Von der DIN 277-1 werden Regeln und Begriffe festgelegt, die zur Ermittlung von Rauminhalten und Grundflächen im Bauwesen dienen. Diese dienen unter anderem dazu, nach ökonomischen Gesichtspunkten zu planen und Bauprojekte durchzuführen, da nur eine Standardisierung auch kostenplanerische Vergleiche möglich macht. Über diese Aufgabe hinaus, nämlich zusammen mit der DIN 276-1 „Kosten im Bauwesen – Teil 1: Hochbau“ eine Grundlage zur Kostenermittlung zu sein, besitzt die DIN 277-1 eine weitere wichtige Stellung. Sie ist durch die in ihr enthaltenen Definitionen und Begriffe zu Grundflächen und Rauminhalten für das Planungs- und Baugeschehen und die Immobilienwirtschaft, schon bevor es zu einer ersten Kostenermittlung kommt, Vergleichsgrundlage für wirtschaftliche Beurteilung, Machbarkeitsstudien etc. Durch das in Bezug Setzen von Planungswerten durch eben in der Norm verankerte Planungskennwerte können nämlich bereits Aussagen zur Wirtschaftlichkeit eines Bauvorhabens getroffen werden. (ebd., S. 450)

Die besondere Bedeutung der DIN 277-1 liegt darüber hinaus darin begründet, dass die aus ihren Regeln ermittelten Grundflächen und Rauminhalte auch für städteplanerische Aktivitäten genutzt werden können. Als Beispiele seien an dieser Stelle die Ermittlung der zulässigen baulichen Nutzung eines Baugrundstückes und die Ermittlung der zulässigen Miet- bzw. Wohnfläche in einem Gebäude genannt (ebd.). Der Anwendungsbereich wird nach DIN 277-1 festgelegt: „Dieses Dokument gilt für die Ermittlung von Grundflächen und Rauminhalten im Hochbau während der Planung, der Bauausführung und der Nutzung von Bauwerken. Es erstreckt sich auf die Grundflächen und Rauminhalte von Bauwerken sowie auf die Grundflächen des Grundstücks“^{2.4}. Die Norm legt „Begriffe, Definitionen, Begriffsinhalte und Regeln für die Ermittlung von Grundflächen und Rauminhalten fest“^{2.4}. Damit wird eine Grundlage geschaffen, die Vergleiche zwischen verschiedenen Bauwerken und Grundstücksflächen jeweils untereinander zulässt. Des weiteren können mit der Norm, wie bereits erwähnt, die „ermittelten Flächen und Rauminhalte [...] auch für andere Zwecke (z. B. die Festlegung der Wohnfläche oder der Mietfläche“^{2.4} herangezogen werden.

^{2.4}DIN 277-1 - Anwendungsbereich

Die Neufassung der Norm im Jahr 2016 stellt einen tiefgreifenden Einschnitt dar (Ruf 2016, S. 457). So wurden unter anderem a)^{2.5} „die DIN 277-1 und die DIN 277-2 [...] gekürzt und zu einem Teil zusammengefasst“; e)^{2.5} „die Norm [...] redaktionell überarbeitet und neu gegliedert“ sowie f)^{2.5} „die Gliederung der Grundflächen des Bauwerks [...] vereinfacht und auf zwei Gliederungsebenen reduziert“. Dies alles ist der Motivation geschuldet, die Norm anwendungsfreundlicher d. h. kürzer und übersichtlicher zu gestalten (ebd.).

BauNVO Boeddinghaus (2000) kommentiert die Baunutzungsverordnung (BauNVO). Die BauNVO ist eng verknüpft mit dem Baugesetzbuch (BauGB). Die BauNVO wendet sich in ihrer Bedeutung für die Bauleitplanung an verschiedene Akteure. In ihren wesentlichen Teilen wendet sie sich an die Gemeinden als die Träger der Bauleitplanung. Hierbei wird zwischen Vorschriften unterschieden, die Bestandteil des Flächennutzungsplans oder des Bebauungsplans sind, und solchen, die darüber hinaus gehen. Auch wendet sich die BauNVO an die Bauaufsichtsbehörden, wenn Abweichungen und Ausnahmeregelungen gefunden werden müssen. Auch der Bauherr und sein Architekt sind von der BauNVO betroffen, wie z.B. bei Baugenehmigungsverfahren und der Bauüberwachung. (ebd., S. 66)

Bei der Aufstellung eines Bebauungsplans sind die Vorschriften der BauNVO vorausgesetzt. Der Bebauungsplan legt die Art und das Maß der baulichen Nutzung sowie die überbaubaren Grundstücksflächen fest. Landesrechtliche Vorschriften ergänzen die BauNVO insbesondere bei der Berechnung von Grundfläche, Geschossfläche und Baumasse. Bevor durch den Bebauungsplan verbindlich festgelegt wird, wie bebaut werden darf und wird, legt der Flächennutzungsplan, meist von der Gemeinde entworfen, vorbereitend und unverbindlich die Grundzüge der Nutzung fest. Es liegt also ein zweistufiges System^{2.6} vor. (ebd., S. 67)

Im Flächennutzungsplan wird auf der obersten Ebene unterteilt in Wohnbauflächen (W), gemischte Bauflächen (M), gewerbliche Bauflächen (G) sowie Sonderbauflächen (S). Dies ist die allgemeine Art der baulichen Nutzung. Die Flächen werden dann bezeichnet als Bauflächen. Nach der besonderen Art der baulichen Nutzung erfolgt die Einteilung detaillierter. Die Flächen werden dann als Baugebiete bezeichnet. Die Wohnbauflächen werden dann unterteilt in Kleinsiedlungsgebiete (WS), reine Wohngebiete (WR), allgemeine Wohngebiete (WA) und besondere Wohngebiete (WB). Gemischte Bauflächen werden unterteilt in Dorfgebiete (MD), Mischgebiete (MI) und Kerngebiete (MK). Gewerbliche Bauflächen werden unterteilt in Gewerbegebiete (GE) und Industriegebiete (GI). Eine allgemeine Sonderbaufläche wird im besonderen Fall zu einem Sondergebiet. (ebd., S. 81)

Der Unterschied zwischen der allgemeinen und besonderen Art liegt darin, dass allgemeine und besondere Arten in einem Flächennutzungsplan aufgeführt werden können. Wird dagegen ein konkreter Bebauungsplan festgelegt, ist es Pflicht, die besondere Art der baulichen Nutzung festzulegen. Dabei

^{2.5}DIN 277-1 - Vorwort/Änderungen

^{2.6}vgl. § 1 Abs. 2 BauGB

können nur die Nutzungsarten zugewiesen, die der Gesetzgeber festgelegt hat. Weitere Nutzungsarten müssen unter dem Punkt Sondergebiet aufgeführt werden. (Boeddinghaus 2000, S. 96)

2.2.3. Bodenrichtwerte

Zimmermann (2010) kommentiert die Immobilienwertermittlungsverordnung (ImmoWertV) und geht dabei auch auf die Ermittlung von Bodenrichtwerten ein. Der Autor macht deutlich, dass die Kenntnis vom Verkehrswert eines Grundstücks eine für viele Entscheidungen wesentliche Vorbedingung darstellt. Er beklagt dabei die fehlende Transparenz des Verfahrens zur Wertermittlung, dass auch noch undurchsichtiger wird, wenn dabei qualitativ schlecht gemachte Sachverständigengutachten auf den Plan treten. Dabei liefert die ImmoWertV zusammen mit dem Baugesetzbuch (BauGB) ein „normativ geregeltes und allgemein verbindliches Recht der Verkehrswertbestimmung“ (ebd., S. V), d. h. es ist durchaus auch für einen nicht sachverständigen Menschen möglich, ein Verkehrswertgutachten zu prüfen. Die Prüfung besteht darin, nachzuvollziehen, ob die anwendbar gewesenen Rechtsvorschriften, hier die ImmoWertV und das BauGB, bei der Erstellung des Verkehrswertgutachten beachtet worden sind. (ebd., S. V f.)

Zunächst müssen die Begriffe „Wert“ und „Bewertung“ bzw. „Wertermittlung“ geklärt werden. Wirtschaftsgüter, zu denen auch Grundstücke zu zählen sind, besitzen einen wie auch immer gearteten Wert. Dieser Wert besteht aber nicht in Währung oder Geld. Eine Methodik, die auf allgemein anerkannter Eingrenzung besteht, zur Umrechnung eines Wertes auf eine bestimmte Menge Währungsgeld wird dann als Bewertung oder Wertermittlung bezeichnet. Dies setzt allerdings einen feststehenden Begriff vom Wert voraus. Dieser ist Ergebnis einer Entwicklung auf gesellschaftlicher Ebene und hat seinen Einfluss bis in die Rechtsprechung genommen, in der dann der Begriff Verkehrswert definiert wurde. (ebd., S. 1 ff.)

Der Bodenrichtwert (BRW) ist ein durchschnittlicher Wert, der auf der Zusammenfassung von einer Vielzahl von einzelnen Kaufpreisen von Grundstücken basiert. Dazu heisst es: „Bodenrichtwerte und sonstige für die Wertermittlung erforderliche Daten sind insbesondere aus der Kaufpreissammlung auf der Grundlage einer ausreichende Zahl geeigneter Kaufpreise unter Berücksichtigung der allgemeinen Wertverhältnisse zu ermitteln. [...]“^{2.7} Damit wird angeordnet, dass die Datengrundlage für die sogenannte Kaufpreissammlung, die von Gutachterausschüssen geführt wird, auf eben einer ausreichenden, repräsentativen Anzahl von einzelnen Kaufpreisen beruht, wobei ausschließlich solche Kaufpreise zu Grunde gelegt werden dürfen, die von Gutachterausschüssen beschlossen worden sind, und nicht etwa sogenannte Erfahrungswerte. Allerdings sind selbst solche von Gutachterausschüssen ermittelten Grundstückspreise kritisch zu sehen, da z. B. fiskalische Interessen von Gemeinden einen Einfluss auf die amtlich ermittelten Verkehrswerte haben können. (ebd., S. 209 ff.)

^{2.7} § 9 (1) ImmoWertV

Zur Ermittlung von Bodenrichtwerten schreibt die ImmoWertV vor: „Bodenrichtwerte sind vorrangig im Vergleichswertverfahren zu ermitteln.[...] Die Bodenrichtwerte sind als ein Betrag in Euro pro Quadratmeter Grundstücksfläche darzustellen“^{2.8}. Dabei kann auch mit deduktiven Verfahren oder vergleichbarem gearbeitet werden, wenn die Anzahl vergleichbarer Preise zu gering ist. Die Einheit der Bodenrichtwerte ist dabei eindeutig vorgeschrieben als €/m². (Zimmermann 2010, S. 219 f.)

Ebenfalls vorgeschrieben ist mittlerweile die elektronische Verarbeitung der Bodenrichtwerte: „Die Bodenrichtwerte sind in automatisierter Form auf der Grundlage der amtlichen Geobasisdaten zu führen“^{2.9}. Dieser Punkt wurde auf Verlangen des Bundesrates in das Gesetz aufgenommen; als Begründung wird die Notwendigkeit angeführt, dass es einen länderübergreifenden Datenaustausch geben soll. Da das Raumbezugssystem für die Geodaten der Bodenrichtwerte dem der amtlichen Liegenschaftskarten entsprechen muss, ist der Informationsaustausch gewährleistet. (ebd., S. 241)

2.2.4. Raumordnung

Schmitz (2004) gibt einen Überblick über die zeitliche Dauerhaftigkeit von Raumordnungsprogrammen. Danach ist die Prognose der Bevölkerungsentwicklung schon immer Grundlage für die Landes- und Regionalplanung sowie die Raumordnung gewesen (ebd., S. 128). Allerdings ist die Übertragbarkeit von großräumigen Prognosen auf kleinräumige Gebiete dabei nicht gegeben, was dazu führt, dass Regionen ihre eigenen Bevölkerungsprognosen durchführen, um auf regional spezifische Gegebenheiten eingehen zu können (ebd., S. 129). Dabei sind Bevölkerungsprognosen auf Bundes- und Länderebene für Vergleiche zwischen den Regionen durchaus geeignet, um eventuelle Defizite einer Region aufzuzeigen (ebd., S. 131). Abgesehen vom räumlichen Faktor ist die Zeit für zwei Aspekte der Raumplanung ausschlaggebend: (ebd., S. 129) zum einen zur Festlegung des Status quo (Gegenwart) und der möglichen zukünftigen Entwicklung; zum anderen für die Definition eines Zielhorizontes für das regionale Entwicklungskonzept. Daher ist es von großer Bedeutung, „bis in welche Phase der prognostizierbaren Bevölkerungsentwicklung die Geltungsdauer eines Planes und seiner Zielhorizonte hineinreicht“ (ebd.). Der Prognosezeitraum der Bevölkerungsentwicklung sollte für eine erfolgreiche Planung ferner in der Zukunft liegen als der Zielhorizont für die Raumplanung. Bisher wird Raumplanung so betrieben, dass der Zielhorizont nicht zeitlich festgelegt wird, so dass eine eigentliche Evaluierung vom Gesamtkonzept her nicht durchgeführt werden kann (ebd., S. 134).

Die wesentlichen Schritte zur Aufstellung eines Regionalplanes (Erarbeitung, Abstimmung, Aufstellung/Beschlussfassung, Genehmigung) dauern in der Regel 3 bis 5 Jahre, in wenigen Fällen auch nur 2 Jahre. Die Geltungsdauer dagegen wird meist auf 10 bis 15 Jahre festgelegt, eine Fortschreibung oder Neufassung geschieht frühestens nach 5 bis 10 Jahren. Davon ausgenommen sind aber z. B. Gebiete, in

^{2.8}§ 10 (1) ImmoWertV

^{2.9}§ 10 (3) ImmoWertV

denen Rohstoffe abgebaut werden oder in denen mit Überschwemmungen zu rechnen ist. Generell ist mit kürzeren Planungszeiträumen eher nicht zu rechnen. Dafür ist der organisatorische Aufwand zu groß. Das jeweilige Landesplanungsgesetz lässt aber Planänderungs- und Zielabweichungsverfahren zu, so dass eine gewisse Flexibilität gewiss vorhanden ist. (Schmitz 2004, S. 130)

Raumplanerisch relevante Elemente von Bevölkerungsprognosen sind die Bevölkerungsentwicklung in der gesamten Bundesrepublik Deutschland sowie den einzelnen Bundesländern, die Altersstruktur, das Erwerbspotential, der Einfluss auf Wirtschaftswachstum und Volkseinkommen, der Strukturwandel, die Zuwanderung, die Konsequenzen für die Siedlungsentwicklung, Arbeitslosigkeit und Mobilität, und regionale Disparitäten (Unterschiede zwischen und innerhalb von Regionen) (ebd., S. 131 ff.). Für die Siedlungsentwicklung kann ein Bevölkerungsrückgang bedeuten, dass zum einem im finanziellen Bereich negative Folgen auftreten. Hier ist dann zu beobachten, dass Grundstückspreise sinken und Produktions- und Infrastrukturanlagen entwertet werden, ja sogar aufgegeben werden müssen. Die Aufgabe von Infrastrukturanlagen und der Rückgang der Bevölkerung haben dann zur Wirkung, dass die verbliebenen Anlagen größere Gebiete versorgen müssen (ebd., S. 133). Der Rückgang an Infrastruktur und höhere Mobilitätskosten führen dann zu einer Flucht vom Land in die Stadt und einer weiteren Verschlechterung der ländlichen Struktur (vgl. Kluge, Libbe und Schramm 2006).

Bisher wurde Regionalplanung vor allem unter den folgenden Aspekten betrieben. Als Planelemente gelten Bevölkerungsrichtwerte, Flächenbedarfswerte, zentrale Orte und Entwicklungsachsen, Siedlungsbereiche, Freiraumfestlegungen. Bevölkerungsprognosen und -richtwerte werden ermittelt und/oder festgelegt. Dies geschieht auch mit Flächenbedarfswerten. Für beide ist der Gültigkeitsbereich bzw. die Auflösung meistens auf regionaler Ebene. Entwicklungsachsen werden aus dem Landesentwicklungsplan übernommen und in den regionalen Bereich projiziert und angepasst. Die Freiraumfestlegungen werden trotz ihres schützenden Charakters meist negativ aufgefasst, da sie die Entwicklung von Gemeinden und Siedlungen einschränken. Demgegenüber steht die Ausweisung neuer Bauflächen, die zwar nicht genutzt werden, für Gemeinden aber als Wachstum angesehen werden (Schmitz 2004, S. 135). Bezüglich der Wohnfunktion von Siedlungen wird eine Bevölkerungsprognose meist nur herangezogen, wenn es um die Entscheidung einer Verteilung von Einwohnern über ein bestimmtes Gebiet geht, beispielsweise bei der Nachverdichtung. Dabei „wird nicht in allen Regionalplänen die Ableitung des Bedarfs [an Wohnfläche] aus regionalen Bevölkerungsvorausschätzungen nachprüfbar nachgewiesen“ (ebd.).

Der Autor kommt insgesamt zu folgendem Fazit (ebd., S. 136 f.): Rechtliche Vorgaben und die nur im Sinne einer räumlichen Planung durchgeführte Regionalplanung beschränken die Steuerungsfunktion der Siedlungsentwicklung. Nur eine wachsende Bevölkerungsentwicklung hat einen Einfluss auf die Planungen. Schrumpft nun die Bevölkerung, hat dies bislang keine Auswirkung auf die Planungen, da der Bevölkerungsrückgang eine neuere, bisher nicht einfließende Entwicklung darstellt. Der bisherige Planungsstil setzt vor allem auf einen Konsens mit den Gemeinden, die auf ihrem Gemeindegebiet

die Akteure der Regionalplanung darstellen. Daher sind Rückplanungen erst zu erwarten, wenn die Gemeinden ihre Rolle bei der Regionalplanung so wahrnehmen, dass die Nahtstelle zwischen örtlicher und überörtlicher Planung geschlossen wird. In der Regionalplanung immer noch unterschätzt werden die rechtlichen Rahmenbedingungen (Eigentums- und Steuerrecht) sowie die Förderpolitik (Wohnungsbauförderung, Pendlerpauschale). Die Steuerungswirkung von Einzelnen (Investoren, Bauträgern) beeinflusst mindestens mittelbar die Planungen. Der Siedlungsdruck wird insgesamt zurückgehen, was auf den anhaltenden Trend zur Suburbanisierung zurückzuführen ist. Die bisherigen Planungsinstrumente werden wohl nicht in Frage gestellt werden, so dass man weiter davon ausgehen kann, dass Regionalplanungen auf eine lange Zeit konzipiert und festgehalten werden.

2.3. Geoinformationssysteme

Bill (2016) beschreibt den Nutzen und die Grundlagen der Geoinformationssysteme (GISE). Zur Bedeutung der GISE schreibt er: „Geo-Informationssysteme (GIS) haben sich zu einem unverzichtbaren Instrumentarium für Wirtschaft, Verwaltung, Wissenschaft und den Bürger entwickelt. GIS ist Bestandteil des täglichen Lebens, nutzt gängige Informationstechnologien, beruht auf Standards und ist in viele Arbeitsabläufe integriert.“ (ebd., S. vii)

Der Einsatz von elektronischer Datenverarbeitung (EDV) in der raumbezogenen Datenerhaltung geht zurück bis in das Jahr 1963. In diesem Jahr veröffentlichte Roger F. Tomlinson in Kanada ein computergestütztes Informationssystem, das in der Lage war, jahrhundertlang verwandtes analoges Material wie Karten, Register oder Bücher mit Hilfe von Rechnern zu verarbeiten und zu speichern. Er führte damit auch den Begriff des *Geographischen Informationssystems* ein. (Walker und Miller 1990)

Die Definition lautet: „Ein GIS ist ein rechnergestütztes System, das aus Hardware, Software und Daten besteht und mit dem sich raumbezogene Problemstellungen in unterschiedlichsten Anwendungsgebieten modellieren und bearbeiten lassen. [...] [Es] bezeichnet sowohl eine Technologie, Produkte als auch Vorhaben zur Bereitstellung und Behandlung von Geoinformationen.“ (Bill 2016, S. 8)

Der im Deutschen Sprachgebrauch oft verwandte Begriff des Geoinformationssystems ist dagegen nicht selbsterklärend, kann doch die einfache Silbe „Geo-“ in vielen Disziplinen anders zugeordnet werden. An sich ist „geo“ aus der Griechischen Sprache entliehen und lässt sich vom Wort *γαια* („gaia“) für Erde ableiten. Damit ist das Bezugsobjekt zwar klar, nicht aber ob es sich um den gesamten Globus, um regionale oder lokale Ausdehnung handelt. In manchen Disziplinen spielt nur die horizontale Ebene eine Rolle, also die Erdoberfläche, in anderen Disziplinen wie z. B. der Meteorologie ist die vertikale Ausdehnung aber ebenso von Bedeutung. (ebd., S. 9 ff.)

Als explizite Bezeichnung im Sinne des Raumbezugs bietet sich daher der Begriff raumbezogenes Informationssystem, engl.: *spatial information system*, eher an. Allerdings setzt sich auch im internationalen Raum der Begriff „Geoinformation“ immer mehr durch (ebd., S. 1). Die in einem Geoinformationssystem gespeicherten und verarbeiteten räumlichen Daten werden als Geodaten, engl.: *spatial data*, bezeichnet (Bellinghoff, Dahmen und Heisig 2015, S. 19).

Die im Kontext der Geoinformation oft verwendeten Begriffe Daten, Information und Wissen sind nicht klar voneinander abgegrenzt. Aus informationstechnischer Sicht bedeutet der Begriff Daten die Bit- oder Byte-weise Abspeicherung von Informationen durch einen Computer. Die Abspeicherung kann dabei strukturiert, semistrukturiert oder unstrukturiert geschehen. Der Begriff Information steht wiederum für Daten in einem speziellen Kontext. Das heißt, dass Daten, die semantisch verknüpft werden, ihnen also eine Bedeutung zugewiesen wird, zu Informationen werden. Den reinen Daten,

die von einem Computer ohne weiteres verarbeitet werden können, fehlen die Interpretationsregeln, um vom Menschen verstanden werden zu können. Als Wissen werden Regeln und Transformationen bezeichnet, mit denen Daten manipuliert werden können. Neues Wissen kann entstehen, wenn altes Wissen auf neue Informationen angewendet wird. Somit kann Information auch als zweckbezogenes Wissen aufgefasst werden. (Bill 2016, S. 1 ff.)

Im Endeffekt ist das verknüpfende Element im Bereich der GIS der Raumbezug, d. h. die Koordinate in x-, y- als auch in z-Richtung (ebd., S. 11). Ein GIS ist also ein spezielles Informationssystem, bei dem Objekte über ihren Raumbezug miteinander in Verbindung stehen. Dass Objekte einen Raumbezug haben, ist dabei nicht selbstverständlich. Dieser kann direkt vorhanden sein (primäre Metrik) oder indirekt hergestellt werden (sekundäre Metrik) (ebd., S. 8). Ist beides nicht der Fall, eignen sich die Daten bzw. Objekte nicht für eine Verarbeitung in einem GIS.

Darüber hinaus ist nicht nur der räumliche Bezug von Geoobjekten untereinander von Bedeutung, sondern ebenso der Bezug auf ein definiertes Koordinatenreferenzsystem. Erst diese Zusatzinformation macht es möglich, Daten aus verschiedenen Quellen mit unterschiedlichen Koordinatensystemen zu verknüpfen. Definiert sind diese Koordinatenreferenzsysteme in der ISO 19111. Die Angabe von Genauigkeiten und Unschärfen vervollständigen dann die Angaben zum Raumbezug (ebd., S. 11).

Als Werkzeug zur Modellierung verfolgt GIS das Ziel, die reale Welt rechnergestützt nachzubilden. Dabei ist allerdings die exakte Abbildung der Wirklichkeit weder erreichbar noch zweckmäßig. Es ist vielmehr darauf aus, in einem speziellen Kontext möglichst exakt die Wirklichkeit wiederzugeben. Durch die Reduktion der Daten in einem Modell ist es unvermeidlich, dass diese verfälscht werden und auch Informationen verloren gehen (ebd., S. 17).

Auf der anderen Seite ermöglicht diese Reduktion im Rahmen der Digitalisierung der Daten diese auf verschiedenste Arten und Weisen miteinander zu verknüpfen und zusammenzustellen (ebd., S. 8). Anstatt die reale Welt möglichst genau wiederzugeben, können verschiedene Sichten auf die Welt gezeigt werden, so wie sie im jeweiligen Anwendungsfall von Nutzen sind.

Das Geoobjekt Hat ein Objekt einen vollständigen Raumbezug, wird es dementsprechend als Geoobjekt oder raumbezogenes/räumliches Objekt, im englischen Sprachraum als *Spatial Feature* oder *Spatial Object* bezeichnet. Darunter wird „eine aus Nutzersicht in einem GIS enthaltene und für seine Arbeit bedeutsame Einheit verstanden, welche mittels Geoinformationen eindeutig beschreibbar, in der Realität identifizierbar und referenzierbar ist“ (ebd., S. 19). Damit ist auch klar, dass ein Geoobjekt eindeutig identifizierbar sein muss, da es die Repräsentation eines realen Objektes, also eines Unikats ist. Um diese 1-zu-1-Beziehung hervorzuheben, benutzt die Sprache der Logik hier sogar den Begriff ein-eindeutig. Ein Geoobjekt enthält sowohl quantitative Informationen (über seine Geometrie) als auch qualitative (in der jeweiligen Thematik) (ebd.). Innerhalb der Software FME, siehe Abschnitt

2.3.1, wird ein Geoobjekt als Feature bezeichnet. Wie bereits erwähnt ist das Geoobjekt der kleinste Baustein eines Geomodells. Ein Geoobjekt kann immer einer bestimmten Klasse zugeordnet werden. Dies wird auch als Objektzugehörigkeit bezeichnet. Ein Objekt erbt dabei bestimmte Eigenschaften der übergeordneten Objektklasse (Bill 2016, S. 21). Es besitzt eine oder mehrere Eigenschaften bzw. Komponenten aus den folgenden Bereichen. Als rein quantitativ gelten geometrische, topologische und temporale Eigenschaften. Die Lage und die Form eines Objektes wird durch seine Geometrie festgelegt. Das gewählte Koordinatenreferenzsystem bestimmt nicht nur die Lage des Objekts sondern auch seine Ausmaße, da innere Orts- und Längenangaben sich in ihren Einheiten immer auf das übergeordnete Koordinatenreferenzsystem beziehen. Geometrien können Punkte, Linien, Flächen, Räume und Kombinationen aus diesen sein. Die Verwendung geometrischer und topologischer Eigenschaften wird durch die Normen ISO 19107 und ISO 19137 international festgelegt. Für die relativen räumlichen Bezüge zwischen Geoobjekten sind die topologischen Merkmale eines Geoobjekts zuständig. Hier werden Nachbarschaftsbeziehungen sowie Knoten, Kanten und Maschen festgehalten, die dazu verwendet werden können, typische Vorgänge wie das Verschneiden (Trimmen) und andere (Mengen-) Operationen zwischen Objektsammlungen durchführen zu können (ebd., S. 19). Temporale Eigenschaften grenzen Zeiträume und -punkte ein, in bzw. zu denen andere Eigenschaften gelten. Verändert sich die Lage eines Objekts mit der Zeit, nicht aber seine innere Geometrie, spricht man von bewegten Objekten. Im Gegensatz zu Geoobjekten mit rein statischen Eigenschaften ermöglicht die temporale Komponente eine dynamische Beschreibung von Objekten (ebd., S. 20). Desweiteren existieren noch thematische Eigenschaften, die die Semantik eines Objekts zusätzlich beschreiben. Diese können je nachdem nominal, ordinal oder metrisch skaliert sein. Metainformationen liefern u.a. die Herkunft und die Art, wie die Daten erfasst worden sind (ebd.). Wichtig ist außerdem noch die graphische Ausprägung, also wie ein Objekt am jeweiligen Endgerät ausgegeben wird. Im Grunde genommen können GIS-Daten um eine Fülle weiterer Informationen ergänzt werden. Im Falle von Video- und Audiodaten ist dann von einem Multimedia-GIS die Rede (ebd., S. 21).

Raumbezogene Datenanalysemethoden Als Analyse wird allgemein die Zerlegung und Zergliederung eines Ganzen in seine Einzelteile bezeichnet, um die inneren Zusammenhänge erkennen zu können. Hierbei wird zwischen qualitativer und quantitativer Analyse unterschieden. Im Gegensatz zur Analyse steht die Synthese für eine Verknüpfung oder Vereinigung von einzelnen Daten. Häufig wird allerdings bei der raumbezogenen Datenanalyse die Synthese mit einbezogen. Dies ergibt einen Regelkreis, der aus Analyse, Synthese, Simulation und Prognose besteht und mehrmals (iterativ) durchlaufen werden muss, bevor ein zufriedenstellendes Ergebnis abgeleitet werden kann. (ebd., S. 453 ff.)

Die Methoden lassen sich einteilen in geometrische, topologische, statistische, Mengen-, temporale sowie weitere komplexere Analysemethoden. Bei den geometrischen Methoden wird abhängig von

der Datengrundlage eine Unterteilung in die Verarbeitungsmethoden nach Vektor- oder Rasterdaten unterschieden (Bill 2016, S. 456).

Als geometrische Methoden seien genannt der Rechtecktest, engl.: *clipping*, mit den Anwendungsbereichen „Punkt gegen Rechteck“, „Kante gegen Rechteck“ sowie „Polygon gegen Rechteck“, der Punkt-im-Polygon-Test, wobei hier die Verfahren zwischen Vektordaten und Rasterdaten stark unterscheiden, die Zonengenerierung, ebenfalls mit starken Unterschieden zwischen Vektor- und Rasterdaten, die Flächenverschneidung (Fläche mit Fläche, Linie mit Fläche, Punkt mit Fläche) sowie die Dreiecksvermaschung (ebd., S. 462 ff.). Der Rechtecktest, bei dem u. a. geprüft werden kann, ob ein Geoobjekt innerhalb eines achsparallelen Rechtecks, das ein weiteres Geoobjekt minimal einschließt, liegt, dient dazu, die Anwendung weiterer komplexerer Methoden auszuschließen (ebd., S. 466 ff.). Der Punkt-im-Polygon-Test, der dazu dient festzustellen, ob ein Punkt sich innerhalb, auf der Grenze oder außerhalb eines Polygons befindet, liefert kein positives Ergebnis, wenn sich der Punkt schon nicht im Rechteck befindet, das das Polygon umschließt, und kann daher übersprungen werden (ebd., S. 462 f.). Die Zonengenerierung, die auch als Pufferbildung bezeichnet wird, dient z. B. der Beantwortung der Fragestellung, ob sich ein Objekt in einem definierten Umkreis oder Abstand von einem weiteren Objekt befindet (ebd., S. 469). Die Flächenverschneidung, engl.: *polygon overlay*, ist die Minimalanforderung an ein GIS und somit auch die wichtigste Analysemethode. Hier werden zwei Eingangsdatenmengen geometrisch überlagert, so dass eine weitere Datenmenge entsteht, die nicht nur die Geometrien, sondern vor allem auch die Sachdaten, also alle nicht-geometrischen Attribute, miteinander vereint (ebd., S. 474 ff.). Die Dreiecksvermaschung dient zur Suche nach den nächsten Nachbarn, engl.: *nearest neighbour*, wobei diese nach der jeweiligen Distanz geordnet werden und dann als 1-, 2- bis k-ter nächster Nachbar bezeichnet werden (ebd., S. 481). Das bekannteste Verfahren bei Vektordaten ist die Delaunay-Triangulation, bei Rasterdaten wird häufig die Mittelachsentransformation angewendet (ebd., S. 482 ff.).

Als Ergänzung zu den geometrischen können die topologischen Methoden aufgefasst werden (ebd., S. 488). Zu ihnen gehören Netzwerkanalysen und Wegeprobleme, als auch die Punktmengentopologie. Im Unterschied zu den geometrischen Methoden, die sich auf Eigenschaften wie Ausdehnung, Form, Größe und Lage beziehen, werden die topologischen Methoden bei Eigenschaften angewendet, die sich beim Drehen, Verzerren, Stauchen usw. nicht ändern (ebd., S. 498 f.). Die Beziehungen eines einzelnen Punktes zu einer Punktmenge können dabei die folgenden sein: der Punkt liegt innerhalb der Punktmenge, engl.: *interior*; der Punkt liegt am Rand der Punktmenge, engl.: *boundary*; der Punkt liegt innerhalb der Punktmenge oder auf dem Rand, engl.: *closure*; der Punkt liegt außerhalb der Punktmenge, engl.: *exterior* (ebd., S. 499). Die topologischen Operatoren sind wiederum eine Kombination aus diesen Beziehungen zwischen Geoobjekten und im folgenden vorgestellt: disjunkt, engl.: *disjoint*, wenn zwei Objekte keine Schnittmenge aufweisen; *meet*, wenn sich zwei Objekte an ihren Außengrenzen berühren; *overlap*, wenn sich zwei Objekte überschneiden, wobei hier die

Überlappung mit Getrenntheit von jener mit Überschneidung unterschieden wird; *contains*, wenn ein Objekt vollständig vom zweiten umschlossen wird; *inside*, bei Umkehrung des vorangegangenen Falls, das zweite Objekt wird vom ersten umschlossen; *covers*, wenn ein Objekt das andere deckt; *covered by*, bei Umkehrung des vorangegangenen Falls; *equal*, wenn beide Objekte miteinander übereinstimmen (Bill 2016, S. 502 ff.).

Die raumbezogenen statistischen Verfahren sind u. a. die Verfahren der räumlichen Interpolation. Weitere statistische Verfahren werden in Abschnitt 2.4 *Statistische Analyse* vorgestellt. Für die räumliche Interpolation stehen die Verfahren der linearen Interpolation, der inversen Gewichtsfunktion, engl.: *inverse distance methode*, und der multiquadratischen Interpolation zur Verfügung (ebd., S. 532 f.). Für die inverse Gewichtsfunktion sind zunächst Nachbarpunkte $t = 1, \dots, T$ zu finden, zu denen Informationen vorliegen, die nun auf einen Punkt, der innerhalb eines Dreiecks, das diese Nachbarpunkte aufspannen, liegt, projiziert werden. Abhängig vom jeweiligen Anwendungsfall wird der Exponent in der Gewichtsfunktion $p_t = \frac{1}{s_t^n}$ mit $n = 1 \dots \infty$ gewählt. Mit der Interpolationsformel werden dann die Informationen aus den Nachbarpunkten abhängig von ihrer Entfernung s_t gewichtet (ebd., S. 532):

$$z = \frac{\sum_{t=1,T} p_t z_t}{\sum_{t=1,T} p_t} \quad (2.5)$$

Bei der invers quadratischen Gewichtsfunktion ist der Exponent dementsprechend $n = 2$.

Mengenmethoden basieren auf der mathematische Grundlage der Mengenlehre. Für deren Beschreibung haben sich die Boole'sche Algebra, die Fuzzy-Mathematik und die relationale Algebra entwickelt (ebd., S. 534). Letztere hat für GIS eine große Bedeutung, da mit Hilfe ihrer Operatoren mit den relationalen Datenbanken, die dem GIS zugrunde liegen, gearbeitet wird. Die Operatoren sind folgende: bei der Projektion wird eine ganze Tabellenspalte ausgewählt; bei der Selektion werden Datensätze aufgrund eines speziellen Attributwertes ausgegeben; bei der Verknüpfung, engl.: *join*, werden zwei Tabellen über ein gemeinsames Attribut verknüpft; beim Produkt wird das kartesische Produkt aus dieser Verknüpfung ausgegeben; bei der Vereinigung wird ein gemeinsamer Datensatz aus der Verknüpfung erzeugt; dagegen enthält die Differenz die Daten, die nur in der ersten Tabelle vorkommen; der Schnitt enthält alle Daten, die in beiden Tabellen gleichzeitig auftreten; bei der Division wird ein zuvor gebildetes Produkt wieder aufgelöst (ebd., S. 548 f.). Ein weiterer Anwendungsbereich von Mengenmethoden sind die Aggregation und die Disaggregation. Bei der Aggregation werden Informationen von einer kleinteiligen Ebene auf eine gröbere Ebene übertragen. Bei der Disaggregation werden im Gegensatz dazu Informationen von einer großskalierten Ebene auf eine feinere Ebene projiziert (ebd., S. 554 ff.). Disaggregation kann z. B. über die bei den raumbezogenen statistischen Verfahren erwähnte räumliche Interpolation erfolgen, wenn Informationen mit einer geringen Dichte auf kleinteilige Gebiete abgebildet werden.

Studien al-Khouri (2003) beschreibt die semantische Zuordnung von 3D-Gebäudedaten zu einer speziellen Nutzungsart. Die Arbeit handelt über die Möglichkeit der automatischen Zuweisung von 3D-Gebäudedaten zur tatsächlichen Nutzung des realen Gebäudes. Hier werden für zwei Datensätze (ländlich, suburban) semantische Regeln angesetzt, die in der Lage sind, aus Gebäudemerkmalen (Traufhöhe, Fürsthöhe, Dachneigung, Anzahl Vollgeschosse etc.) die Nutzung eines Gebäudes als Wohnhaus, Garage, Schuppen, Scheune etc. zu bestimmen (ebd., S. 73 ff.). Diese Methodik ist allerdings für Wohngebiete entwickelt worden und nicht ohne weiteres auf andere Siedlungsformen übertragbar.

Die Arbeit von al-Khouri (ebd.) beinhaltet eine wissensbasierte Methode zur Erkennung verschiedener Gebäudetypen. Hier werden GIS-Daten mit Bauvorschriften, Dachmerkmalen und anderen aus GIS stammenden Daten kombiniert. Die Zuordnung von Gebäuden findet zu einer vorher definierten Klasse statt. Im einem weiteren Schritt werden zusätzlich Unterklassen in Haupt- und Nebengebäude gebildet. (ebd., S. 1)

„Die Modellierung [...] erfolgt durch Regeln, die den Zusammenhang zwischen Gebäudetypen und deren Merkmalen darstellen“ (ebd.). Diese Regeln basieren auf Fakten (Bauvorschriften) und Beobachtung der zu untersuchenden Gebiete und sind daher nicht allgemein gültig und nicht ohne weitere Überlegungen auf andere Gebiete übertragbar. Es werden Sicherheitsfaktoren beaufschlagt, die einerseits durch zwei verschiedene Theorien untermauert werden. Diese sind eine bedingte Wahrscheinlichkeitstheorie und die Dempster-Shafer-Evidenztheorie.

Gebäude sind anthropogenen, künstlichen Ursprungs. Sie sind vielfältig und komplex und daher trotz ihrer Künstlichkeit schwer zu erfassen und zu kategorisieren (ebd., S. 5). Ihre Eigenschaften variieren in Form, Struktur, Richtung und Flächenmaterialien, um nur einige physische Merkmale zu nennen (ebd., S. 6). Sie können von verschiedenen Disziplinen (Architektur, Ökonomie, Soziologie, Technologie, Jurisprudenz) auf unterschiedliche Art und Weise in ihrer Erscheinung und Funktion gedeutet und interpretiert werden (ebd., S. 5). Die 3D-Gebäudedaten, die vor allem aus Laserscanning und digitalen Geländemodellen (DGM) stammen, allein liefern nur die geometrischen Merkmale eines Gebäudes. Demgegenüber liefern klassische Kartengrundlagen und moderne GIS-Daten weitere Informationen über die Geometrie hinaus, sie enthalten also implizite Informationen. GIS speichern also sowohl geometrische als auch semantische Daten. Sie dienen „der Erfassung, Verarbeitung, Analyse und Präsentation raumbezogener Daten“ (vgl. Bill 2016, S. 6). Dabei ist vor allem bei solchen anthropogenen Objekten, die einer ständigen Veränderung bzw. auch einem Auftauchen und Verschwinden unterliegen, die ständige Aktualisierung der Daten notwendig, damit diese konsistent sind (al-Khouri 2003, S. 5).

Als Informationsfusion versteht man die Zusammenführung von unterschiedlichen Daten aus diversen Quellen. Die dabei neu entstandene Stufe an Wissen kann eine genauere Interpretation sowie eine

höhere Beschreibungsebene ermöglichen (al-Khouri 2003, S. 11). Für die Zusammenführung von Daten müssen diese jeweils mindestens eine gemeinsame Eigenschaft besitzen, aufgrund derer die Daten miteinander verknüpft werden können.

Als Objektmodelle werden geordnete Sammlungen von Daten bezeichnet. Ihre Merkmale können geometrischer, physischer, struktureller oder semantischer Natur sein. Sie können durch Mengenoperationen oder Kausalketten miteinander verknüpft sein. Um die formale Logik auf diese Daten anwenden zu können, müssen diese streng geordnet sein und mit einem einmaligen Identifizierungsmerkmal versehen sein (ebd., S. 12).

2.3.1. Verwendete Software

Für die Aufbereitung der Geodaten kommen zwei Programme zum Einsatz, die sich gegenseitig in ihrem Funktionsumfang ergänzen. Als reines Geoinformationssystem zur Aufbereitung und visuellen Darstellung von Kartenmaterial kommt die Software *QGIS* zum Einsatz. Für die Verarbeitung der Daten und die Manipulation der Geobjekte wird die Software *FeatureManipulationEngine* verwendet.

QGIS Die Software QGIS, frühere Bezeichnung *QuantumGIS*, ist ein Geoinformationssystem (GIS), das unter der GNU Public License (GPL) herausgegeben wird. Das bedeutet, dass der Quellcode einsehbar ist und entsprechend der Lizenz verändert werden darf und das Programm kostenfrei aus dem Internet heruntergeladen, genutzt und weitergegeben werden kann (QGIS Project 2017, S. 5). Die genutzte Version ist *2.14.11-Essen* unter der *GNU GENERAL PUBLIC LICENSE Version 2* vom Juni 1991.

FME Die Software *FeatureManipulationEngine*, kurz FME, der Firma *SafeSoftware Inc.* aus Vancouver, Kanada, wurde ursprünglich dazu entwickelt, um eine Möglichkeit des Datenaustausches zwischen Geoinformationssystemen und Software zur computergestützten Entwicklung, engl.: *computer aided design*, kurz CAD, zu schaffen. Die für die vorliegende Arbeit verwendete Software ist die FME Database Edition in der Version *FME(R) 2017.1.0.0 (20170731 - Build 17539 - linux-x64)*, für die von der Firma *SafeSoftware Inc.* auf Anfrage eine Volllizenz für ein Jahr kostenfrei im Rahmen des *Safe Software Academic Grant program*^{2.10} zur Verfügung gestellt wurde. FME wird auch als Spatial ETL Software bezeichnet, wobei ETL für *extract* (extrahieren), *transform* (transformieren) und *load* (laden) steht. Die Software ist in der Lage, mittlerweile über 350 verschiedene Dateiformate zu lesen. Darunter befinden sich Formate mit Geometrien wie Raster-, Vektor und Punktwolkendaten als auch solche mit thematischem Inhalt wie Sachdaten. Dabei werden mit Hilfe eines sogenannten Readers die

^{2.10} E-Mail-Korrespondenz vom 07.09.2017

Datenquellen eingelesen und in ein internes Format umgewandelt. Die Daten können nun innerhalb der Workbench von FME beliebig bearbeitet werden. Dazu gehören die Möglichkeiten der Anreicherung und Ausdünnung von Datensätzen, der Veränderung der Struktur und der Transformation dieser. Schließlich können die Daten durch sogenannte Writer in Ausgabedateien oder Datenbanken geschrieben werden. (Bellinghoff, Dahmen und Heisig 2015, S. 19)

Das FME-Feature-Modell ist die Basis für die Verarbeitung von Geodaten. Ein FME-Feature ist im Grunde ein Geoobjekt, dass sich durch folgende Eigenschaften auszeichnet. Der Featuretyp ist die Zusammenfassung mehrerer Features aufgrund einer oder mehrerer verbindender Eigenschaften. Dies wird auch Objektgruppierung genannt. Als Featuretyp aufgefasst werden beispielsweise Tabellen (aus Datenbanken), Layer (aus Geoinformationssystemen) und Level (aus digitalen Geländemodellen). Die Eigenschaften eines Geoobjekts werden als Attribute bezeichnet. Hier unterscheidet FME drei verschiedene Arten solcher Objekteigenschaften. Vom jeweiligen Benutzer frei definierbare Attribute bzw. solche, die vom jeweiligen Datenmodell vorgegeben sind, werden als Benutzerattribute bezeichnet. Ihre Namen können ebenso frei gewählt werden, vorausgesetzt, die Namen sind noch nicht vergeben oder werden nicht durch die Benutzung eines speziellen Dateiformates begrenzt. Das interne Datenmodell von FME speichert Informationen z.B. über Farben und Formatierungen in sogenannten FME-Attributen. Diese Attribute sind mit einem Präfix gekennzeichnet, z.B. `fme_color`. Spezifische Merkmale aus verschiedenen Datenformaten werden in sog. Formatattributen abgelegt. Dies sind beispielsweise Linienstile oder Füllmuster. Abhängig vom Datenformat besitzen sie auch ein kennzeichnendes Präfix, z.B. `gml_id` als Identifikationsnummer eines Objektes in der *Geography Markup Language*, kurz GML. Die Anzahl und Art der Formatattribute variiert stark zwischen den verschiedenen Formaten. Die räumliche Lage sowie die Gestalt eines FME-Features werden durch die Eigenschaften der Geometrie definiert. Dabei können keine, eine oder mehrere Geometrien möglich sein, die von einfachen Typen wie Punkt, Linie und Fläche bis zu komplexen Typen wie Aggregaten, 3D-Geometrien, Raster und Punktwolken reichen. Geometrien können Namen (`geometry name`), Eigenschaften (`traits`) sowie Bemaßungsinformationen (`measures`) haben. Zuletzt kann jedes Feature auch Informationen enthalten, auf welches Koordinatensystem es sich bezieht. Üblicherweise ist dies aber eine Eigenschaft des Featuretyps; ist also ein verbindendes Element einer Gruppe von Features. (ebd., S. 64 ff.)

Die Prozesse zur Konvertierung und Transformation geschehen in der zentralen Komponente von FME, der FME Workbench (ebd., S. 73). Auf ihrer Arbeitsfläche, engl.: *canvas*, werden die Elemente eines Projektes angeordnet und zu einem Prozessfluss verbunden. Dies geschieht grafisch, so dass der Anwender intuitiv mit ihnen arbeiten kann, ohne tiefes informationstechnisches Wissen besitzen zu müssen (ebd., S. 75). Ein Projekt in der FME-Workbench wird auch als Workspace bezeichnet. Die Elemente eines Projektes sind die Reader, Transformer und Writer. Wie schon erwähnt, werden mit Hilfe der Reader die Daten eingelesen. Sie werden typischerweise links auf der Arbeitsfläche

positioniert. Auf der rechten Seite befinden sich die Writer, die die Daten am Ende des Prozesses in Ausgabedateien schreiben. In der Mitte der Arbeitsfläche stehen die Transformer, mit denen die Daten manipuliert und transformiert werden (Bellinghoff, Dahmen und Heisig 2015, S. 78). Hinsichtlich ihrer Funktionsweise werden die Transformer unterteilt in solche, die ein Feature nach dem anderen sequentiell verarbeiten, und solche, die eine Gruppe von Features gleichzeitig verarbeiten (ebd., S. 92 f.). Bei den Feature-basierten Transformern wird jedes Feature unabhängig von den anderen verarbeitet. Die Reihenfolge der Features wird dabei eingehalten. Als Beispiel sei der Transformer *AreaCalculator* genannt, der es ermöglicht, für jedes Feature seine eigene Fläche zu berechnen. Bei den gruppenbasierten Transformern läuft eine Gruppe von Features auf den Eingang des Transformers auf und wird dann gleichzeitig manipuliert. Die Reihenfolge der Features wird dabei nicht eingehalten; das System der Anordnung ist unbekannt, da es von der Firma *SafeSoftware Inc.* intern gehalten wird. Als Beispiel für einen gruppenbasierten Transformer sei hier der *AreaBuilder* genannt, mit dem mehrere Linien zu einem Polygon vereinigt werden können (ebd., S. 93). Weitere Transformer werden in den Beschreibungen zu den erstellten Workspaces vorgestellt.

2.4. Statistische Analyse

Dieckmann (2012) macht darauf aufmerksam, dass für die Herstellung und Belegung eines (statistischen) Zusammenhangs einige Grundlagen zuerst geklärt und festgelegt werden müssen. Nach dem Hempel-Oppenheim-Schema ^{2.11} werden empirisch zutreffende Hypothesen benötigt, um geeignete Maßnahmen zu treffen. (s. ebd., S. 186) Für die Erstellung eines Forschungsproblems ist eine abhängige Variable, deren unabhängige Variable noch gefunden werden muss, einzugrenzen und zu definieren. Dazu sagt Dieckmann (ebd., S. 187), dass „ein Forschungsproblem klar definieren heißt z. B. bei einer hypothesenprüfenden Untersuchung, dass die Forschungshypothese (abhängige und unabhängige Variable[n]) präzise angegeben werden kann“. Da die Frage nicht nur lautet, ob es einen Zusammenhang gibt, auch nicht nur ob der Zusammenhang positiv oder negativ ist, sondern welchen Zusammenhang es gibt, d. h. wie geartet dieser Zusammenhang ist, ist mindestens eine Ordinalskala notwendig. Da aber „die Abstände zwischen den Skalenwerten nicht sinnvoll interpretierbar“ (ebd., S. 287) sind, wird mindestens eine Intervallskala benötigt. Diese wiederum hat die Eigenschaft, dass „wegen der willkürlichen Festlegung des Nullpunkts (oder eines anderen Anfangswertes) (ebd., S. 288) eine Vergleichbarkeit nicht möglich ist. Abhilfe schafft die Nutzung einer Verhältnisskala, da sich hier „ein ‚natürlicher Nullpunkt‘ festlegen“ (ebd., S. 289) lässt.

„Sind unabhängige und abhängige Variable *quantitativ* (d. h. die Variablen haben mindestens *Intervallskalenniveau*), dann kann eventuell die genaue Form des Zusammenhangs durch eine mathematische Funktion dargestellt werden.“ (ebd., S. 132). Im einfachsten Fall ist dies ein linearer Zusammenhang der Form $y = a \cdot x + b$, mit y als abhängige und x als unabhängige Variable.

Explorative Untersuchungen Diese werden durchgeführt, wenn das zu betrachtende Feld noch relativ unerforscht ist. Häufig wird vor einer Hauptsudie eine explorative Phase durchgeführt, die auch als Vorstudie oder Pre-Test angesehen werden kann. Bevorzugt werden qualitative Studien, sowie Experteninterviews und unstrukturierte Beobachtungen, (vgl. ebd., S. 33 f.) .

Somit wird durch explorative Studien ein ‚Gefühl‘ für ein neues Feld erst entwickelt: „Das Forschungsziel ist hier nicht die Prüfung von Hypothesen, sondern deren Entwicklung anhand des empirischen Materials.“ (ebd., S. 188)

Deskriptive Untersuchungen Im Mittelpunkt steht hier die Schätzung von Häufigkeiten, Anteilen, Durchschnittswerten und andere Verteilungsmerkmalen. Dabei sind „repräsentative“ Studien nötig, wie z. B. Zufallsstichproben oder sogar Totalerhebungen. Das Ziel ist nur die Beschreibung (!) und

^{2.11} „nach C.G. Hempel und P. Oppenheim benanntes Muster wissenschaftlicher Erklärungen: Einen realwissenschaftlichen Sachverhalt zu erklären heißt, ihn auf logisch-deduktivem Wege (Deduktion) aus Gesetzesaussagen und Anwendungsbedingungen abzuleiten.“ (Springer Gabler Verlag 2015)

Diagnose, (s. Dieckmann 2012, S. 34). Die Fragestellung lautet: *bei genau welcher Population sollen welche Merkmale geschützt werden?* (ebd., S. 187)

Prüfung von Hypothesen und Theorien Sie ist nötig, wenn es zum selben Sachverhalt mehrere Hypothesen gibt, die zu widersprechenden Prognosen führen. Bei der empirischen Prüfung von Hypothesen und Theorien müssen Unsicherheiten reduziert und Fehlerquellen, wie Verzerrungen z. B. Selbstselektion^{2.12} und Versuchsleitereffekte, unter Kontrolle gebracht werden, (s. ebd., S. 37).

Die Fragestellung lautet hier: *kann die Forschungshypothese (abhängige und unabhängige Variable(n)) präzise angegeben werden?* vgl. ebd., S. 187

Evaluationsstudien Diese werden häufig in der Grundlagenforschung angewendet. Hierbei wird die Wirksamkeit einer Maßnahme anhand von Erfolgskriterien abgeschätzt. Dabei werden auch eventuelle Nebenwirkungen untersucht. vgl. ebd., S. 37 f.

2.4.1. Univariate Häufigkeitsverteilungen

Kühnel und Krebs (2007) empfehlen vor der Durchführung von bi- oder multivariaten Analysen von Daten die Durchführung einer univariaten Datenanalyse, um die Verteilung der Ausprägung (Realisation) einer einzelnen Variablen zu betrachten, damit ein erster Eindruck gewonnen werden kann (ebd., S. 41). Eine Häufigkeitstabelle dient dazu, die Ausprägungen einer Variablen ihrer Häufigkeit nach zu ordnen. Dabei werden diese der Größe nach sortiert. Um die Interpretation zu erleichtern, werden sogenannte relative Häufigkeiten verwendet. Die relative Häufigkeit p_k einer Ausprägung x_k der Variablen X berechnet sich zu:

$$p_k = \frac{n_k}{n} \quad (2.6)$$

wobei n_k die absolute Häufigkeit der Realisation x_k und n die Gesamtfallzahl darstellt. Dabei kann die Gesamtfallzahl alle Fälle umfassen oder nur die gültigen Werte, was je nach Qualität der Datenbasis stark von einander abweichen kann. Die relative Häufigkeit wird oft auch als Prozentwert angegeben. Die relativen Häufigkeiten der einzelnen Ausprägungen können auch addiert werden. Man spricht dabei von Kumulierung. Werden alle Werte addiert, ergibt sich 1, 0 bzw. 100%. Dieser Zusammenhang kann der Prüfung der Berechnungen dienen. (ebd., S. 43 ff.)

^{2.12}Solche Selektionsverzerrungen sind durch strikt experimentelle Untersuchungsanordnungen zu vermeiden, (vgl. Dieckmann 2012, S. 65)

Die empirische Verteilungsfunktion ist eine mathematische Funktion, die für jeden Wert den Anteil der Fälle angibt, deren Ausprägung kleiner oder gleich diesem Wert sind. Der Funktionswert entspricht der Spalte mit den kumulierten Anteilen in der Häufigkeitstabelle. Die grafische Darstellung der empirischen Verteilungsfunktion entspricht einer treppenförmigen Kurve (Kühnel und Krebs 2007, S. 50). Hat eine Variable wenige Realisationen, so können die Häufigkeiten in einer Tabelle übersichtlich dargestellt werden. Bei metrischen Werten ist dies aber selten der Fall, so dass eine Einteilung sinnvoll ist (ebd., S. 43 ff.).

Klassenbildung Als Klasse bezeichnet man eine Zusammenfassung, also eine Gruppierung, von Ausprägungen einer Variablen. Eine Klasse besitzt eine untere und eine obere Grenze sowie eine Klassenmitte bzw. einen Klassenmittelpunkt. Der Abstand zwischen der unteren und der oberen Grenze einer Klasse wird als Klassenbreite oder Klassenintervall bezeichnet.

Um sicherzugehen, dass bei einer Gruppierung so wenige Informationen wie möglich verloren gehen, muss bei der Bildung von Klassen auf folgende Regeln geachtet werden:

1. Klassen dürfen sich nicht überschneiden, damit eine Ausprägung einer Variable nur einer Klasse zugeordnet werden kann.
2. Die Klassen dürfen untereinander keine Lücken aufweisen, so dass die Ausprägung einer Variable keiner der Klassen zugeordnet werden könnte.
3. Die Breite der Klasse sollte gleich groß sein: gleich große Intervalle. Es kann aber vorkommen, dass z.B. Ausreißer am oberen und unteren Rand der Ausprägungen vorliegen und es somit sinnvoll ist, hier an den Rändern die Klassenbreite zu vergrößern.
4. Bei metrischen Daten müssen die Klassengrenzen exakt wiedergegeben werden. Dies erfolgt nach den Regeln über das Runden von Zahlen. Zum Beispiel gilt für die Zahl 1, dass die exakte untere Grenze bei 0,5 und die exakte obere Grenze bei $< 1,5$ liegt.

Die obere Grenze wird mit o_k , die untere Grenze mit u_k bezeichnet. k steht für die Klasse. Für die Klassenmitte gilt:

$$m_k = \frac{u_k + o_k}{2} \quad (2.7)$$

Dazu ergänzt Holland und Scharnbacher (2006, S. 17), dass die Übersicht bei großen Datenmengen schnell verloren gehen kann und deshalb eine Einteilung in Klassen die Übersicht erhöht. Allerdings gehen bei jeder Gruppierung Informationen verloren, je größer die gewählte Klassenbreite, desto mehr Informationen gehen verloren, desto mehr steigt allerdings die Übersichtlichkeit.

Quantile Durch ein Quantil wird eine Häufigkeitsverteilungen in zwei Hälfte geteilt. Dabei wird die relative Häufigkeit im unteren Teilbereich als Quantilanteil α bezeichnet. Die Grenze liegt dann am Quantilwert Q_α . Das heisst, dass Q_α der kleinste Wert ist, für den der Anteil der Fälle mit gleicher oder kleinerer Ausprägung größer oder gleich diesem Wert ist. Bei ungruppierten Daten lassen sich die Quantilwerte direkt aus den kumulierten Häufigkeiten bilden. Bei gruppierten Daten erfolgt die Ermittlung über eine Summenkurve, auch Summenfunktion genannt. Hierzu ist allerdings metrisches Skalenniveau Voraussetzung. (Kühnel und Krebs 2007, S. 51 ff.)

Grafische Darstellung Durch die grafische Darstellung von Häufigkeitsverteilungen können Ergebnisse übersichtlich und kompakt dargestellt und miteinander verglichen werden. Es existieren eine Vielzahl an Darstellungsarten, im Folgenden werden nur einige genannt. Stabdiagramme, Histogramme und Kern-Dichte-Schätzer ermöglichen einen umfassenden Überblick über die gesamte Häufigkeitsverteilung. Letztere beiden setzen metrisches Skalenniveau voraus. Für den Vergleich von verschiedenen Verteilungen bieten sich Box-Plots an, die sich auf einige zentrale Aspekte der empirischen Häufigkeitsverteilung konzentrieren und zur visuellen Darstellung von Kennwerten univariater Verteilungen dienen. (ebd., S. 56 f.)

Kennwerte Die Kennwerte von univariaten Verteilungen, auch als Parameter, Maßzahlen oder Maße von Verteilungen benannt, haben den Zweck, die Eigenschaften einer Verteilung im Gegensatz zu Häufigkeitstabellen und den grafischen Darstellungsarten in möglichst wenigen Werten zusammenzufassen. Dabei geht es vor allem um die Ermittlung typischer Kennwerte, die zum einen das Maß der zentralen Tendenz, auch als Lagemaße bezeichnet (Modus, Median, Mittelwert), und zum anderen die Unterschiedlichkeit in der Verteilung der Realisation einer Variablen bezüglich der Streuung (Spannweite, Variation, Varianz, Standardabweichnung) abbilden. In einigen Fällen ist es von Nutzen, auch die Schiefe oder die Steilheit einer Verteilung zu bestimmen. (ebd., S. 67)

Die am häufigsten auftretende Realisation einer Verteilung ist der Modus, auch als Modalwert oder engl. *mode* bezeichnet. Bei ungruppierten Daten entspricht dieser der Realisation mit der größten absoluten bzw. relativen Häufigkeit. Bei gruppierten Daten entspricht der Modus dem Wert der Klassenmitte der Klasse mit der größten absoluten wie relativen Häufigkeit (ebd., S. 67 f.). Alle Skalenniveaus (nominal, ordinal, metrisch) erlauben die Berechnung des Modus. Dies hat aber auf der anderen Seite den Nachteil, dass bei einer alleinigen Betrachtung dieses Kennwertes Unterschiede zwischen verschiedenen Verteilungen übersehen werden können und dass Verteilungen, die eigentlich mehrere Gipfel (multimodal) aufweisen, fälschlicherweise als unimodal gewertet werden können (ebd., S. 80).

Wenn die Werte einer Verteilung der Größe nach geordnet in genau zwei gleich große Teile bezogen auf die Anzahl der Realisationen in der Mitte geteilt wird, entspricht der Wert, der eben in der

Mitte liegt, dem Median \tilde{x} . Bei ungerader Fallzahl erhält man diesen Wert direkt, er ist gleich dem Quantilwert $Q_{50\%}$ des 50 %-Quantils. Dies gilt für ungruppierte Daten uneingeschränkt. Bei gerader Fallzahl muss die Position des Medians interpoliert werden. (Kühnel und Krebs 2007, S. 67 f.)

Bei gruppierten Daten muss auch hier interpoliert werden. Hier entspricht der Median dem interpolierten 50 %-Quantil (ebd., S. 71). Für die Berechnung des Medians müssen die Werte geordnet sein. Dies trifft auf ordinal- und metrisch-skalierte Messniveaus zu. Der Median ist sehr robust gegenüber Ausreißern, hat aber auch den Nachteil, dass, wenn nur er zur Bewertung einer Verteilung herangezogen wird, Charakteristika einer Verteilung unentdeckt bleiben. (ebd., S. 80 f.)

Das arithmetische Mittel ist ein Durchschnittswert, der den Schwerpunkt einer Verteilung angibt (ebd., S. 67). Daher wird er auch als Mittelwert oder Durchschnittswert, engl. *mean*, bezeichnet. Er trägt das Symbol \bar{x} und ist die Summe aller Ausprägungen einer Variablen bezogen auf die Fallzahl (ebd., S. 72). Daher enthält dieser Kennwert auch die meisten Informationen über eine Verteilung im Vergleich zu Modus und Median. Allerdings ist er dadurch auch anfälliger für Ausreißer (ebd., S. 81 f.).

Im Gegensatz zu den bisher genannten Kennwerten, die allesamt ein Maß für die zentrale Tendenz darstellen, sind die folgenden Kennwerte Streuungsmaße, die einen Überblick über die Heterogenität einer Verteilung erlauben. Die typischen Kennwerte sind Spannweite, Variation, Varianz und Standardabweichung (ebd., S. 83 ff.).

Die Spannweite, engl. *range*, ist der Abstand zwischen dem kleinsten und dem größten Wert einer Verteilung, also die Differenz zwischen Minimum und Maximum. Sie ist allerdings sehr instabil gegenüber den Extremwerten, deswegen wird oft der Quartilabstand zu einer besseren Vergleichbarkeit herangezogen. Beiden gemeinsam ist die Tatsache, dass sie nur auf zwei Werten aus der Verteilung beruhen. Daher werden noch weitere Kennwerte benötigt. (ebd., S. 84 ff.)

Die Variation, auch als mittelwertbereinigte Quadratsumme, engl. *sum of squares*, bezeichnet, ist eine Möglichkeit, zu verhindern, dass sich Abweichungen vom Mittelwert aufheben und einen falschen Eindruck vermitteln (ebd., S. 88). Für eine Variable X wird die Standardabweichung mit SS_X symbolisiert und berechnet zu:

$$SS_X = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \quad (2.8)$$

Wenn die Variation auf die Anzahl der Fälle bezogen wird, erhält man die Varianz. Als Symbol wird s_X^2 verwendet. Da dies zur Folge hat, dass Zahlen und Einheiten quadriert vorliegen, wird mit der Standardabweichung gearbeitet. Diese ist die Quadratwurzel der Varianz und wird mit s_X symbolisiert. Schließlich kann diese Standardabweichung noch auf den Mittelwert bezogen werden, um

zu berücksichtigen, dass Verteilungen mit großen Mittelwerten oft große Standardabweichungen besitzen. Dieser Variationskoeffizient V_X als Standardabweichung bezogen auf den Anteil am Mittelwert wird vor allem für den Vergleich von Verteilungen mit sehr unterschiedlichen Mittelwerten herangezogen. (Kühnel und Krebs 2007, S. 89 ff.)

2.4.2. Regression und Korrelation

Das lineare Regressionsmodell, mit dem ein Zusammenhang zwischen zwei Variablen verifiziert oder falsifiziert werden kann, beruht auf der Annahme, dass die Regressionskurve in einer Population eine Gerade ist. Dabei liegen die Mittelwerte $\mu_{Y=f(X)}$ der abhängigen Variablen $Y = f(X)$ auf dieser Regressionsgeraden. Bezogen auf die gesamte Population wird eine Variable $M_{Y=f(X)}$ eingeführt, die alle Mittelwerte umfasst. Diese ist im linearen Regressionsmodell eine lineare Funktion der Realisationen der unabhängigen Variable X . Mit der Regressionskonstante α und dem Regressionsgewicht β kann dann die folgende Gleichung aufgestellt werden, die die bedingten Populationsmittelwerte der abhängigen Variablen Y als Funktion von X beschreibt (ebd., S. 384 f.):

$$M_{Y=f(X)} = \alpha \cdot \beta + X \quad (2.9)$$

Die Parameter α und β werden zusammen als Regressionskoeffizienten bezeichnet. Sie können über die Methode der kleinsten Quadrate bestimmt werden. Dazu werden die Koeffizienten so abgeschätzt, dass die Summe der quadrierten Residuen, also die Abstände zwischen der Regressionsgeraden und den Datenpunkten, minimal ist. Diese Schätzer a und b werden dazu verwendet, die Regressionsgleichung aufzustellen. Diese lautet dann (ebd., S. 392):

$$\hat{Y} = a + b \cdot X \quad (2.10)$$

Die Regressionskonstante a bzw. α gibt den Schnittpunkt der Regressionsgeraden mit der Ordinate an. Sie ist damit die Angabe über die Lage der Regressionsgeraden. Aus ihrem Wert lässt sich in der Praxis selten eine Interpretation ableiten. Dagegen hat das Regressionsgewicht b bzw. β die Bedeutung, dass es den Effekt der unabhängigen Variable auf die abhängige direkt angibt. Je höher sein Betrag, desto stärker reagieren die Vorhersagewerte der abhängigen Variable auf eine Änderung in der unabhängigen Variable. (ebd., S. 394 ff.)

2.4.3. Clusteranalyse

Das Ziel einer Clusteranalyse besteht darin, aus einer Vielzahl von Objekten diejenigen zu Gruppen zusammenzufassen, die die größte Ähnlichkeit untereinander, jedoch eine große Unterschiedlichkeit zwischen den Gruppen aufweisen. Dazu schreiben Backhaus et al. (2016, S. 455), dass diese Gruppen „im Hinblick auf die betrachteten Eigenschaften oder Merkmale als möglichst homogen zu bezeichnen sind“ und „gleichzeitig [...] untereinander eine möglichst große Heterogenität aufweisen“. Diese Gruppen werden als Cluster bezeichnet. Im Bereich der multivariaten Datenanalyse ist die Clusteranalyse den explorativen Untersuchungsmethoden zuzuordnen, da die Anzahl und die Gestalt der Gruppen vor ihrer Bildung noch unbekannt sind. Allen Clusterbildungsverfahren ist gemein, dass sie alle Eigenschaften der betrachteten Objekte gleichzeitig zur Gruppenbildung heranziehen, wobei es natürlich eine Auswahl geben muss, so dass nur solche Eigenschaften herangezogen werden, die theoretisch einen Sinn ergeben, also als relevant für den zu untersuchenden Sachverhalt angesehen werden können (vgl. ebd., S. 510). Dies hat zur Folge, dass vor einer Clusteranalyse die zu betrachtenden Merkmalsausprägungen der Variablen normalisiert werden müssen, worauf später noch genauer eingegangen wird.

Zur Clusteranalyse gehören die folgenden Schritte (ebd., S. 456 ff.). Der erste Schritt besteht in der Wahl des Proximitätsmaßes, mit dem die Ähnlichkeit, bzw. des Distanzmaßes, mit dem die Unähnlichkeit zwischen den Merkmalen von Objekten gemessen wird. Dazu ist es zunächst nötig, die Rohdatenmatrix, die die zu betrachtenden Variablen enthält, in eine Ähnlichkeitsmatrix bzw. Distanzmatrix umzuwandeln. Diese ist dann immer eine quadratische Matrix, da hier alle Maße der einzelnen Objekte untereinander aufgeführt sind. Unterschieden werden zwei Arten (ebd., S. 458): Ähnlichkeiten werden durch Ähnlichkeitsmaße dargestellt; je größer der Wert des Ähnlichkeitsmaßes ist, desto ähnlicher sind sich zwei Objekte. Auf der anderen Seite stehen die Distanzmaße, die die Unähnlichkeit messen; wenn die Objekte identisch sind, ist das Distanzmaß gleich null. Proximitätsmaße gibt es sowohl für Variablen mit metrischem, nominalem als auch mit binärem Intervallskalenniveau. Da in der vorliegenden Arbeit hauptsächlich metrisch skalierte Variablen zum Einsatz kommen, wird nur auf diese Proximitätsmaße eingegangen. Als Distanzmaße werden von Backhaus et al. (ebd., S. 459) Kosinus und Pearson-Korrelation genannt. Als Distanzmaße führen die Autoren die (quadrierte) euklidische Distanz, die Minkowski-Metrik, die Block-Metrik und die Tschebyscheff-Metrik auf. Die (quadrierte) euklidische Distanz wird dabei zu den am weitesten verbreiteten Distanzmaßen in der Empirie gezählt (ebd., S. 469). Die quadrierte euklidische Distanz hat die Eigenschaft, dass durch die Quadrierung große Distanzwerte einen größeren Einfluss auf das Ergebnis haben, als kleinere Werte. Die quadrierte euklidische Distanz hat im n -dimensionalen Raum für zwei Vektoren/Variablen

x, y die folgende Form:

$$d^2(x, y) = (x_1 - y_1)^2 + \cdots + (x_n - y_n)^2 = \sum_{i=1}^n (x_i - y_i)^2 \quad (2.11)$$

Die einfache euklidische Distanz erhält man durch Ziehung der Quadratwurzel:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.12)$$

Hier sei angemerkt, dass für den Fall $n = 2$ der allseits bekannte Satz des Pythagoras vorliegt. Im Folgenden kommt die einfache euklidische Distanz zum Einsatz; die quadrierte euklidische Distanz sei hier nur der Vollständigkeit halber genannt. Mit der euklidischen Distanz lässt sich nun die Distanzmatrix aufbauen, was für den weiteren Schritt Voraussetzung ist.

Der zweite Schritt der Clusteranalyse besteht in der Auswahl des Fusionierungsalgorithmus (Backhaus et al. 2016, S. 475 ff.). Hier werden zunächst monothetische, bei denen die Gruppenbildung anhand einer einzelnen Variable geschieht, und polythetische Verfahren unterschieden. Der große Vorteil der polythetischen Verfahren besteht darin, dass die Gruppierung nach mehreren Merkmalsausprägungen erfolgen kann. Daher wird auch ein solches Verfahren bevorzugt. Ist die Anzahl der Cluster bereits bekannt oder kann festgesetzt werden, bieten sich partitionierende Verfahren an, die Objekte zwischen den einzelnen, schon definierten Gruppen, austauschen, bis ein Optimum erreicht ist. Kann am Anfang die Anzahl der Cluster nicht festgelegt werden, ist hingegen ein hierarchisches Verfahren notwendig, in dessen Verlauf erst die Anzahl der Cluster errechnet wird. Auch bei den hierarchischen Verfahren muss wiederum unterschieden werden: befinden sich alle Objekte am Anfang der Clusterbildung in einer Gruppe, und wird diese Gruppe in immer kleiner werdende Gruppen unterteilt, liegt eine divisive Berechnungsmethode vor; befindet sich am Anfang der Berechnung jedes Objekt in seiner eigenen Gruppe, d. h. die Anzahl der Gruppen entspricht der Anzahl der Objekte, und werden diese Gruppen Schritt für Schritt zusammengefasst, spricht man von agglomerativen Algorithmen. Ihnen kommt in der Praxis eine größere Bedeutung zu (ebd., S. 476). Allen agglomerativen Verfahren ist folgender Ablauf gemein (ebd., S. 478):

1. Jedes Objekt befindet sich in der eigenen Gruppe; stellt eine eigene Gruppe dar
2. Für alle Objekte wird paarweise das Proximitätsmaß berechnet.
3. Die beiden Cluster mit der größten Ähnlichkeit werden zusammengefasst.
4. Die Ähnlichkeitsmatrix wird daraufhin reduziert und neu berechnet.

-
5. Es werden die Schritte 2. und 3. solange durchlaufen, bis alle Objekte in einer einzigen Gruppe sind.

Um möglichst homogene Gruppen zu bilden, bietet sich das Ward-Verfahren an. Untersuchungen haben ergeben, dass es in der Lage ist, „wahre Gruppen“ zusammenzufassen und damit eine „wahre Gruppierung“ zu erreichen (Backhaus et al. 2016, S. 494). Dementsprechend besitzt es auch eine weite Verbreitung in der praktischen Anwendung (ebd., S. 484). Hier werden nur Gruppen mit geringen Unterschieden vereinigt, damit ein vorgegebenes Heterogenitätsmaß nicht überschritten wird. Als Heterogenitätsmaß kann als Varianzkriterium die Fehlerquadratsumme oder die schon oben beschriebene quadratische euklidische Distanz genutzt werden. Bei Zusammenfassung von zwei Objekten bzw. Gruppen entspricht dann der Wert in der Distanzmatrix (nach der quadratischen euklidischen Distanz) genau dem doppelten Wert der Fehlerquadratsumme. Backhaus et al. (ebd., S. 494) empfehlen außerdem, dass mit Hilfe des Single-Linkage-Verfahrens geprüft werden soll, ob Ausreißer vorhanden sind, um diese vor der eigentlichen Clusteranalyse zu entfernen.

Als letzter Schritt der Clusteranalyse steht die Bestimmung der (optimalen) Clusteranzahl an (ebd., S. 494 ff.). Hierfür bietet sich das sogenannte Elbow-Kriterium an. Durch einen einfachen Scree-Plot, bei dem die Anzahl der Cluster aus den einzelnen Fusionierungsschritten auf der Abszisse und die Summe der Fehlerquadrate auf der Ordinate aufgetragen wird, kann, wenn vorhanden, die optimale Clusterzahl durch den Knick im Graphen abgelesen werden. Allerdings ist dies kein mathematisch erbrachtes Ergebnis, sondern eine subjektive Einschätzung. Es gibt aber auch statistische Kriterien, die die optimale Anzahl an Clustern bei einer hierarchischen Clusteranalyse anzeigen. Diese werden in der Literatur als *stopping rules* bezeichnet (ebd., S. 496). Als bestes Verfahren geben Backhaus et al. (ebd.) das Verfahren nach Calinski/Harabasz an, das bei Tests in der Lage war, in 90 % aller Fälle die „wahre Gruppenstruktur“ zu erkennen. Das Verfahren vergleicht die Streuung innerhalb der Gruppen W (engl. *within*, innerhalb) mit der Streuung zwischen den Gruppen B (engl. *between*, dazwischen). Auf eine weitere mathematische Beschreibung wird an dieser Stelle verzichtet.

2.4.4. Verwendete Software

Für die automatisierten statistischen Berechnungen wird auf die Software R zurückgegriffen. Die Software liegt in der Version 3.4.3 vom 30.11.2017 unter den Bedingungen der *GNU General Public License* vor. Diese lässt sich in die Software FME mit Hilfe eines speziellen Transformers, dem *RCaller* einbinden. Diesem Transformer wird sozusagen ein R-Script übergeben, welches dann von einem integrierten Interpreter abgearbeitet wird.

R ist eine Sprache und Umgebung für statistische Berechnungen und Grafiken. Es ist ein GNU-Projekt, das der S-Sprache und Umgebung ähnlich ist, die von John Chambers und Kollegen entwickelt wurde. R kann als eine andere Implementierung von S betrachtet werden. Es gibt einige

wichtige Unterschiede, aber viel Programmcode, der für S geschrieben wurde, läuft unverändert unter R. R bietet eine breite Palette von statistischen (lineare und nichtlineare Modellierung, klassische statistische Tests, Zeitreihenanalyse, Klassifikation, Clustering, ...) und grafischen Techniken und ist sehr erweiterbar. Die S-Sprache ist oft das Mittel der Wahl für die statistische Methodologieforschung, und R stellt dabei einen Open-Source-Weg dar. (R Core Team 2017)

3. Datengrundlage

Die gesamte in dieser Arbeit verwendete Datengrundlage ist frei verfügbar. Auch aus diesem Grund wurde das Land Nordrhein-Westfalen ausgesucht, da das landeseigene Portal für GIS-Daten diese kostenfrei und unkompliziert zur Verfügung stellt. Im Falle der Daten für die Bevölkerungsdichten, die von der Bundesrepublik Deutschland frei zur Verfügung gestellt werden, trifft dies ebenfalls zu. In diesem Kapitel werden die verwendeten Datenquellen mit ihren Eigenschaften vorgestellt.

3.1. Bestandsdatenauszug

Der Bestandsdatenauszug des Landes Nordrhein-Westfalen ist Teil des AFIS-ALKIS-ATKIS (AAA)-Vorhabens. Dieses wird im nächsten Abschnitt kurz beschrieben. Es besteht aus drei Geoinformationssystemen, wobei in dieser Arbeit das Amtliches Topographisch-Kartographisches Informationssystem (ALKIS) von besonderem Interesse ist, das es die Daten zu den Grundstücken bzw. Flurstücken enthält, um zusammen mit den 3D-Gebäudedaten siedlungsstrukturelle Kennzahlen wie die Grundflächenzahl (GRZ) und die Geschossflächenzahl (GFZ) zu errechnen. Die AdV beschreibt im ALKIS-Objektartenkatalog die sehr umfassende Sammlung der verschiedenen Typen von Objekten, siehe Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) (2008).

Im Gegensatz zu dem bei GIS gebräuchlichen Ausdruck des Geoobjekts, siehe Abschnitt *Das Geoobjekt* in Unterkapitel 2.3 *Geoinformationssysteme*, spricht man beim AAA-Vorhaben von Fachobjekten. Dies ist dadurch begründet, dass nicht nur reale Gegenstände, sondern vor allem auch Sachverhalte abgebildet werden (Bill 2016, S. 650).

Flurstücke Wie bereits erwähnt beinhaltet dieser Katalog Daten zu den Flurstücken. Die Definition ist dabei die folgende: „Flurstück‘ ist ein Teil der Erdoberfläche, der von einer im Liegenschaftskataster festgelegten Grenzlinie umschlossen und mit einer Nummer bezeichnet ist. Es ist die Buchungseinheit des Liegenschaftskatasters“ (Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) 2008, S. 43). Abgesehen von den geometrischen Daten besitzt ein Objekt des Typs AX_Flurstueck weitere Attribute bzw. Relationen, die in Tabelle 3.1 aufgelistet sind. Die Daten in dieser Tabelle wurden dem ALKIS-Objektartenkatalog (ebd., S. 43 ff.) entnommen.

Durch die Attribute `istGebucht`, `zeigtAuf` und `weistAuf` lassen sich Flurstücke zu größeren Flur- bzw. Grundstücken zusammenfassen. Die Daten werden im Bestandsdatenauszug bewusst getrennt

geführt, Stichwort: Elementarobjekt. Es macht für die vorliegende Arbeit Sinn, diese zusammenzufassen, damit z. B. mehrere Flurstücke, die im Grundbuch als eines hinterlegt sind, auch als solches behandelt werden.

Bezeichnung	Definition
zustaendigeStelle	'Flurstück' wird verwaltet von 'Dienststelle'. Diese Attributart wird nur dann belegt, wenn eine fachliche Zuständigkeit über eine Gemarkung bzw. Gemarkungsteil/Flur nicht abgebildet werden kann. Die Attributart enthält den Dienststellenschlüssel der Stelle, die fachlich für ein Flurstück zuständig ist.
istGebucht	Ein (oder mehrere) Flurstück(e) ist (sind) unter genau einer Buchungsstelle gebucht. Bei Anteilsbuchungen ist dies nur dann möglich, wenn ein fiktives Buchungsblatt angelegt wird. Wird ein fiktives Buchun[g]sblatt verwendet, ist die Kardinalität dieser Attributart 1..1.
zeigtAuf	'Flurstück' zeigt auf 'Lagebezeichnung ohne Hausnummer'.
weistAuf	'Flurstück' weist auf 'Lagebezeichnung mit Hausnummer'.
gehörtAnteiligZu	'Flurstück' gehört anteilig zu 'Flurstück'. Die Relationsart kommt nur vor bei Flurstücken, die eine Relation zu einer Buchungsstelle mit einer der Buchungsarten Anliegerweg, Anliegergraben oder Anliegerwasserlauf aufweisen.

Tabelle 3.1: Attribute und Relationen der Objektart AX_Flurstueck, Kennung: 11001, Auszug aus Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) 2008, S. 43 ff.

Eng verknüpft, aber ausgelagert, weil sie auch für historische Flurstücke gelten, sind die Kerndaten der Flurstücke. Einige Attribute sind in Tabelle 3.2 aufgelistet. Das Attribut **gemarkung** macht es möglich, die Flurstücke einer Gemarkung zuzuordnen, bzw. zu identifizieren, welcher Gemarkung sie zugeordnet sind. Gleiches gilt für das Attribut **flurnummer**, was eine Einteilung in die jeweilige Flur zulässt. Über das Attribut **gemeindezugehoerigkeit**, das aus einer 8-stelligen Ziffer besteht und unterteilt ist in einen Wert für das Bundesland (Ziffern 1 und 2), den Regierungsbezirk (Ziffer 3), den Kreis (Ziffern 4 und 5) und die Gemeinde (Ziffern 6 bis 8), lässt sich das Flurstück eindeutig einer Gemeinde zuweisen.

Bezeichnung	Definition
gemarkung	'Gemarkung' enthält die Eigenschaften aus dem Datentyp AX_Gemarkung_Schlüssel: 'land' und 'gemarkungsnummer'.
flurnummer	'Flurnummer' ist die von der Katasterbehörde zur eindeutigen Bezeichnung vergebene Nummer einer Flur, die eine Gruppe von zusammenhängenden Flurstücken innerhalb einer Gemarkung umfasst.
gemeindezugehoerigkeit	'Gemeindezugehörigkeit' enthält das Gemeindekennzeichen zur Zuordnung der Flu[r]stücksdaten zu einer Gemeinde.

Tabelle 3.2: Attribute und Relationen der Objektart AX_Flurstueck_Kerndaten, Kennung: 11004, Auszug aus Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) 2008, S. 55 ff.

Das Attribut **gemeindezugehoerigkeit** enthält aber nur den Verweis auf eine weitere Objektklas-

se, in der dann der Gemeindeschlüssel gespeichert ist. Tabelle 3.3 listet die Attribute dieser Klasse auf. Die Definition der Klasse ist: „'Gemeindekennzeichen' ist die vom Statistischen Bundesamt veröffentlichte Schlüsselnummer des kommunalen Gebietes (Stadt-, Landgemeinde, gemeindefreies Gebiet)” (Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) 2008, S. 456).

Bezeichnung	Definition
land	Bundesland
regierungsbezirk	Regierungsbezirk. Diese Attributart ist optional, da nicht in allen Ländern Regierungsbezirke vorkommen.
kreis	Kreis
gemeinde	Gemeinde
gemeindeteil	Gemeindeteil

Tabelle 3.3: Attribute und Relationen der Objektart AX_Gemeindekennzeichen, Kennung: 73014, Auszug aus Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) 2008, S. 456 f.

Gebäude Die Gebäude aus dem Bestandsdatenauszug sind zugleich Grundlage für die Erstellung der dreidimensionalen Gebäudedaten der Bundesländer gewesen. Insofern besitzen beide Datenquellen die gleiche Identifikationsnummer für die Gebäude, was eine automatisierte Verarbeitung und Verknüpfung der beiden Datenquellen vereinfacht, da sie auf der selben Quelle basieren. Nach dem ALKIS-Objektartenkatalog ist eine Gebäude wie folgt definiert: „'Gebäude' ist ein dauerhaft errichtetes Bauwerk, dessen Nachweis wegen seiner Bedeutung als Liegenschaft erforderlich ist sowie dem Zweck der Basisinformation des Liegenschaftskatasters dient.” (ebd., S. 190 ff.)

In Tab. 3.4 sind einige Attribute des Objekttyps AX_Gebaeude aufgelistet. Das Attribut `gebaeudefunktion` kann diverse Werte annehmen. Dabei sind verschiedene Klassen verfügbar: Werten von 1000 bis unter 2000 sind Gebäude zugeordnet, die im allgemein gefassten Bereich Wohnen liegen. Werten von 2000 bis unter 3000 sind solche zugeordnet, die eine Funktion in Wirtschaft oder Gewerbe erfüllen. Werten von 3000 bis unter 4000 sind öffentliche Gebäude zugeordnet. Den Wert 9998 erhält ein Gebäude, das „Nach Quellenlage nicht zu spezifizieren <ist>” (ebd., S. 198), was bedeutet, dass keine Aussage über die Wertart gemacht werden kann.

Weitere Objektklassen Des weiteren gibt es noch die Objektklasse `AX_LagebezeichnungMitHausnummer`. Diese gibt die Möglichkeit zur Verknüpfung von Flurstücken und Gebäuden. Sie muss aber vorhanden sein, damit dies erreicht werden kann. So lauten die Konsistenzbedingungen (ebd., S. 66): „Die Relation zum Objekt 'AX_Georeferenzierte Gebäudeadresse' muss nur dann gebildet werden, wenn die Relation zu einem Objekt 'AX_Gebäude' existiert und wenn 'AX_GeoreferenzierteGebaeude' dauerhaft

Bezeichnung	Definition
gebaeudfunktion	'Gebäudefunktion' ist die zum Zeitpunkt der Erhebung vorherrschend funktionale Bedeutung des Gebäudes (Dominanzprinzip).
weitereGebaueudfunktion	'Weitere Gebäudefunktion' ist die Funktion, die ein Gebäude neben der dominierenden Gebäudefunktion hat.
bauweise	'Bauweise' ist die Beschreibung der Art der Bauweise.

Tabelle 3.4: Attribute und Relationen der Objektart AX_Gebaeude Auszug aus Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) 2008, S. 190 ff.

im ALKIS-Bestand geführt wird. Bei Änderungen des Objekts 'AX_LagebezeichnungMitHausnummer' muss stets auch das Objekt 'AX_Georeferenzierte Gebäudeadresse' entsprechend fortgeführt werden." Die Verknüpfung zwischen Gebäuden und Flurstücken muss daher mit umfassenderen Mitteln erfolgen.

Wohngebiete und Mischgebiete Die Objektart AX_Wohnbauflaeche repräsentiert Siedlungsflächen, die zum Wohnen genutzt werden. Daher lautet die Definition nach Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) (2008, S. 218): „Wohnbaufläche' ist eine baulich geprägte Fläche einschließlich der mit ihr im Zusammenhang stehenden Freiflächen (z.B. Vorgärten, Ziergärten, Zufahrten, Stellplätze und Hofraumflächen), die ausschließlich oder vorwiegend dem Wohnen dient". Die Objektart AX_FlaecheGemischterNutzung repräsentiert Siedlungsflächen mit gemischter Nutzung, also auch zum Wohnen. Dabei ist die Definition: „Fläche gemischter Nutzung' ist eine bebaute Fläche einschließlich der mit ihr im Zusammenhang stehenden Freifläche (Hofraumfläche, Hausgarten), auf der keine Art der baulichen Nutzung vorherrscht. Solche Flächen sind insbesondere ländlich-dörflich geprägte Flächen mit land- und forstwirtschaftlichen Betrieben, Wohngebäuden u.a. sowie städtisch geprägte Kerngebiete mit Handelsbetrieben und zentralen Einrichtungen für die Wirtschaft und die Verwaltung." (ebd., S. 236)

Mit Hilfe dieser beiden Datensätze lassen sich die Flurstücke bestimmen, die zum Wohnen genutzt werden. Da sich in den angegebenen Flächen auch Verkehrswege befinden, die sich nicht zum Wohnen eignen, müssen diese ebenfalls eingelesen werden.

Verkehrsflächen Die Objekte der Art AX_Strassenverkehr „umfasst alle für die bauliche Anlage Straße erforderlichen sowie dem Straßenverkehr dienenden bebauten und unbebauten Flächen" (ebd., S. 247). Es gibt noch weitere Datensätze zu Spazierwegen, Bahnstrecken, Flugverkehr etc., die aber insgesamt weniger häufig sind und somit nicht eingelesen werden. Durch die Einbeziehung der Verkehrsflächen können diejenigen Flurstücke, die sich zwar in einem Wohn- oder Mischgebiet befinden,

aber keine Baugrundstücke darstellen, sondern dem Verkehr zur Verfügung stehen, herausgefiltert werden.

AAA-Richtlinie Das AAA-Vorhaben besteht aus drei Basisinformationssystemen. Diese sind die folgenden:

- Amtliches Festpunkt-Informationssystem (AFIS)
- Amtliches Liegenschaftskataster-Informationssystem (ATKIS) sowie
- Amtliches Topographisch-Kartographisches Informationssystem (ALKIS),

wobei letzteres die Vereinigung der bisher getrennt geführten Automatisierten Liegenschaftskarte (ALK) und des Automatisierten Liegenschaftsbuches (ALB) darstellt (Bill 2016, S. 646). Hierzulande ist die Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) für die Standardisierung und den Aufbau dieser Informationssysteme zuständig. Die drei Systeme werden als gemeinsames Anwendungsschema in der GeoInfoDok dokumentiert und beschrieben (ebd., S. 647). Damit werden die einzelnen Kataloge und ihre Objektarten miteinander in Beziehung gesetzt, was einen Datenaustausch und Verknüpfungen möglich macht. Durch das AFIS-ALKIS-ATKIS-Referenzmodell werden die Datenstrukturen vorgegeben (ebd., S. 648). Das Referenzmodell ist unterteilt in drei Ebenen.

Auf der ersten Ebene, der Regelungsebene, werden die Produkte des Modells definiert. Objektartenkataloge (OK) und Signaturenkataloge (SK) regeln Inhalte, Strukturen und Herstellungsvorschriften (ebd.). Die Objektartenkataloge basieren auf den bisherigen Katalogen, wurden aber soweit wie möglich semantisch angepasst. Gleichzeitig ist ihre Struktur durch die ISO-Norm 19110 *Feature Cataloguing Methodology* definiert, welche für diesen Anwendungsfall um weitere benötigte Inhalte erweitert wurde (ebd.). Ein Objektartenkatalog wird automatisch aus dem zugrunde liegenden UML-Modell^{3.1} gewonnen. Die Signaturenkataloge dienen zur Vorgabe von Darstellungen der Objektarten bezogen auf einen Maßstab. Dies umfasst Vorschriften zur Abbildung von Fixpunkten, des Liegenschaftskatasters und der Topographie, Vorschriften zur Bildung von Präsentations- und Kartengeometrieobjekten, Vorschriften zur Darstellung und kartografischen Gestaltung der Objekte sowie Vorschriften zur Ausgestaltung von analogen Auszügen (ebd., S. 649). Signaturenkataloge sind für die vorliegende Arbeit als nachrangig zu betrachten, da das Interesse nicht in einer amtlich exakten Darstellung von Material besteht.

Auf der zweiten Ebene, der Produktionsebene, werden die Vorlagen für die digitale Erfassung von Daten definiert (ebd., S. 649 f.). Die analogen Bestandsdaten wie Karten, Luftbilder etc. werden mit zusätzlichen Daten angereichert und in die Systeme abgespeichert. Dies geschieht nach den im

^{3.1}Unified Modeling Language (UML), deutsch: vereinheitlichte Modellierungssprache

Objektartenkatalog festgelegten Regeln. Hier wird auch festgelegt, welche Kartengeometrieobjekte in Karten mit einem spezifischen Maßstab abgebildet werden. Dies hat den Vorteil, dass die Ausgabe von Kartenmaterial aus den Systemen grundlegend automatisiert verläuft. Das Ziel ist dabei, dass die Bestandsdaten vollständig mit der Beschreibung der einzelnen Fachobjekte, ihrer Darstellung in Text und Karten im System hinterlegt werden.

Die dritte Ebene, die Kommunikationsebene, ist die für den Anwender sichtbare Ebene. Auf ihr werden dem Nutzer die Daten zur Verfügung gestellt. Die Art und Weise, in der die Daten zur Verfügung gestellt werden, ist durch die Normbasierte Austauschschnittstelle (NAS) vorgegeben (Bill 2016, S. 650). Digitale Daten können damit als Vektorformate wie Geography Markup Language (GML) oder als Rasterformate wie TIFF^{3.2} bezogen werden.

Im AAA-Basisschema sind die Fachobjekte auf der untersten Ebene als nicht weiter unterteilbar definiert. Man spricht hier von Elementarobjekten. Diese sind (ebd., S. 651):

- Raumbezogene Elementarobjekte (AA_REO)
- Raumbezogene Elementarobjekte in drei Dimensionen (AA_REO_3D)
- Nicht raumbezogene Elementarobjekte (AA_NREO)
- Zusammengesetzte Objekte (AA_ZUSO)
- Punktmengenobjekte (AA_PMO)

Format und Verfügbarkeit Die Daten aus dem AAA-Vorhaben sind bundesweit verfügbar. Im Fall von Nordrhein-Westfalen ist der Zugriff über ein landeseigenes Portal (www.opengeodata.nrw.de) möglich. Hier können die Daten aus dem Bestandsdatenauszug unkompliziert heruntergeladen werden. Viele andere Bundesländer bieten nur den Zugriff auf Daten über einen vorher gewählten Kartenausschnitt an, im Fall von Nordrhein-Westfalen können alle Daten über direkte Download-Links komplett geladen werden. Das Format der Daten entspricht der INSPIRE-Richtlinie der EU. Als Abkömmling der Markup-Language-Sprache XML^{3.3} ist INSPIRE XML in der Lage, verschiedene Daten aufzunehmen, vorausgesetzt sie sind vorher definiert. So beinhaltet es auch die Möglichkeit, GML Daten zu speichern. INSPIRE steht für *INfrastucture for Spatial Information in Europe* (ebd., S. 234). INSPIRE ist ein Geodateninfrastrukturvorhaben auf europäischer Ebene, das durch die am 15. Mai 2007 in Kraft getretene EU-Richtlinie (2007/2/EG) gesetzlich verankert wurde und somit für alle Mitgliedstaaten der Europäischen Union Gültigkeit besitzt. Es werden Daten gespeichert, die sich auf einen Bereich beziehen, in dem ein Mitgliedstaat der EU Hoheitsbefugnisse besitzt, und die von einer Behörde erstellt worden sind. Die digitale Form ist Voraussetzung für die Speicherung. Die Umsetzung

^{3.2} Tagged Image File Format (vgl. Bill 2016, S. 638)

^{3.3} Extensible Markup Language (XML), deutsch: Erweiterbare Auszeichnungssprache

dieses EU-Gesetzes geschieht auf nationaler Ebene in der Bundesrepublik Deutschland durch die Geodateninfrastruktur Deutschland (GDI-DE). Es handelt sich hierbei über eine ressortübergreifende Maßnahme von Bund, Länder und Kommunen (Bill 2016, S. 237). Die Verfügbarkeit solcher offener Verwaltungsdaten wird in Zukunft wohl noch erleichtert, u. a. weil die Politik sich schon 2010 in der *Dresdner Vereinbarung zum 5. Nationalen IT-Gipfel* einer zentralen für alle Bürger zugänglichen *Open-Government-Plattform* bekannt hat. Dies wird unterstrichen durch den *Nationalen Aktionsplan der Bundesregierung zur Umsetzung der Open-Data-Charta der G8* von 2014, der bis 2020 umgesetzt werden soll (ebd., S. 262 f.).

Das verwendete Koordinatensystem ist die Transverse-Mercator-Projektion für den 32. Abschnitt in der nördlichen Hemispähre. Die Abkürzung lautet ETRS89.UTM-32N oder EPSG:25832. Der Stand der verwendeten Daten für Nordrhein-Westfalen ist vom 01.07.2017.

3.2. 3D-Gebäudedaten

Auf Initiative des Landes Nordrhein-Westfalen, das im Bereich Opendeodata u. a. mit seinem oben erwähnten Portal eine Vorreiterrolle spielt, wurde zusammen mit 70 anderen Institutionen der Standard CityGML geschaffen (ebd., S. 255). Mit digitalen Stadtmodellen kann eine Stadt bzw. Stadtteile und einzelne Gebäude virtualisierbar und virtuell begehbar gemacht werden. Aber auch die Verknüpfung auf semantischer Ebene und die Auswertung der Geometriedaten kann dadurch erfolgen. Dabei besitzt der Standard CityGML die weiteste Verbreitung in diesem Bereich (ebd.).

Verfügbarkeit und Kosten Die Verfügbarkeit der Daten ist unter anderem abhängig davon, ob ein Bundesland die Daten selbst verwaltet, oder die Verwaltung und Verteilung der AdV überlässt. Davon sind auch die Kosten abhängig, denn die AdV erhebt eine Gebühr auf die Daten; einzelne Bundesländer erheben Kosten unabhängig von der AdV-Gebührenrichtlinie, stellen bestimmten Benutzerkreisen ihre Daten mit Einschränkungen zur Verfügung oder haben einen Ansatz der freien Verfügbarkeit (Opendata) gewählt.

Das 3D-Gebäudemodelle im Blockmodell ist die unterste Detailstufe, engl: *Level of Detail (LoD)*. Hier erscheint ein Gebäude als „Klötzchen“, d. h. es wird mit einem Flachdach dargestellt. Verschiedene Dachformen sind hier nicht erkennbar. Mit Stand von 31.12.2016 gibt die AdV an, dass in allen 16 Bundesländer der BRD 3D-Gebäude in der Stufe LoD1 zu 100 % verfügbar seien. Dabei ist die Aktualität der Daten unterschiedlich. Bayern, Mecklenburg-Vorpommern, Nordrhein-Westfalen, Sachsen-Anhalt, Schleswig-Holstein und Thüringen haben einen Datenbestand aus dem Jahr 2016, einige Bundesländer liegen wenige Jahre zurück; nur in wenigen Bundesländern ist der Datenbestand veraltet. Dies liegt zum einen daran, dass die Erfassung von 3D-Gebäudedaten erst seit kurzem praktiziert wird und zum anderen daran, dass die wiederholte Erfassung schnell erfolgen kann.

Abgesehen von der hohen Verfügbarkeit und der Aktualität sind die Daten jedoch meist kostenpflichtig. Die AdV berechnet nach der AdV-Gebührenrichtlinie (AdV-GR) 3D-Gebäudedaten für die Detailstufe LoD1 mit EURO 0,27 pro Objekt mit einer Maximalgebühr von EURO 225.000,- . Für die Detailstufe LoD2 werden für ein Objekt EURO 0,65 erhoben, mit einer Maximalgebühr von EURO 540.000,- (AdV 2016, S. 15). Allerdings gibt es einen Mengenrabatt nach einem gestaffeltem System, siehe Tabelle 3.5, wonach mit steigender Objektanzahl der Preis sinkt, bzw. mit einem sinkenden Faktor geringer als 1 multipliziert wird, der wiederum auf den vorangegangenen Faktor aufgeschlagen wird.

Nr.	Objektanzahl	Faktor	Hinweis
(1)	1 bis einschließlich 1.000	1	
(2)	1.001 bis einschließlich 10.000	0,5	zusätzlich zu (1)
(3)	10.001 bis einschließlich 100.000	0,25	zusätzlich zu (1) und (2)
(4)	100.001 bis einschließlich 1.000.000	0,125	zusätzlich zu (1) bis (3)
(5)	1.000.001 bis einschließlich 10.000.000	0,0625	zusätzlich zu (1) bis (4)
(6)	über 10.000.000	0,03125	zusätzlich zu (1) bis (5)

Tabelle 3.5: Ermäßigungsfaktoren nach Objektanzahl nach AdV 2016, S. 3

Für die Bundesländer Berlin, Hamburg, Nordrhein-Westfalen und Thüringen werden die Daten frei und ohne vorherige Lizenzierung bereit gestellt. Die Verfügbarkeiten der 3D-Gebäudedaten sind in Tabelle 3.6 zusammengefasst. Die Hansestadt Bremen erhebt ein Gebühr von EURO 50,- für Qualifikationsforschungsarbeiten, andernfalls erhöhen sich die Kosten auf mehr als EURO 100,- für die flächendeckenden LoD1-Daten^{3.4}. Ein flächendeckendes 3D-Stadtmodell für ganz Sachsen würde im Rahmen einer Studienarbeit ca. EURO 650,- kosten, regulär ca. EURO 65.000,-^{3.5}. Das Land Niedersachsen stellt seine Daten zu den in der *Kostenordnung für das amtliche Vermessungswesen (KOVerM) vom 25. März 2017*^{3.6} angegebenen Preisen zu Verfügung, die denen der AdV-Gebührenrichtlinie entsprechen. Das Land Sachsen-Anhalt ist bereit, 3D-Gebäudedaten kostenfrei zur Verfügung zu stellen, wenn ein offizieller Antrag von Seiten der Hochschule gestellt wird. Werden diese veröffentlicht, ist allerdings eine Lizenzierung notwendig^{3.7}. Auch für Thüringen stehen die Daten kostenfrei zur Verfügung. Für die Datenbereitstellung wird eine Aufwandsgebühr verlangt, die entfällt, wenn die Daten selbstständig heruntergeladen werden. Das Land Mecklenburg-Vorpommern stellt ebenfalls die Daten kostenfrei in das Internet, wobei für den Download der Daten eine Lizenzierung nötig ist^{3.8}. In Rheinland-Pfalz wird zwar nach der AdV-Gebührenrichtlinie abgerechnet, jedoch erhält man als Student außerhalb des Bundeslandes einen Rabatt von 80% mit einem Mindestpreis von EURO

^{3.4}E-Mail-Auskunft Geodatenservice Hansestadt Bremen am 01.06.2017

^{3.5}E-Mail-Auskunft Staatsbetrieb Geoinformation und Vermessung Sachsen vom 07.06.2017

^{3.6}E-Mail-Auskunft Landesamt für Geoinformation und Landesvermessung Niedersachsen (LGLN) vom 22.12.2017

^{3.7}E-Mail-Auskunft Landesamt für Vermessung und Geoinformation Sachsen-Anhalt am 08.06.2017

^{3.8}E-Mail-Auskunft Landesamt für innere Verwaltung, Fachbereich Geodatenbereitstellung vom 09.06.2017

50,-^{3,9}. Die Antwort des Landes Rheinland-Pfalz lieferte wichtige Nebeninformationen unter anderem, dass der „Gebäudegrundriss [...] dem amtlichen Liegenschaftskatasterinformationssystem (ALKIS) entnommen“ wird. Es werden somit „nur Gebäude dargestellt, die zum Zeitpunkt des Produktionsbeginns Bestandteil in ALKIS waren“. Alle weiteren Bundesländer erheben eine Gebühr nach der AdV-Gebührenrichtlinie.

Die zweite Detailstufe enthält neben den in die Vertikale extrudierten Gebäudegrundrissen auch die Informationen über die Dachformen. Diese wurden von einem Algorithmus aus standardisierten Dachformen entlang der tatsächlichen Ausrichtung des Dachfirstes ausgerichtet. Bei der höheren Detailstufe LoD2 liegt die Zeitschiene bis zum Erreichen der bundesweiten Festlegungen in den Bundesländern zum Teil erst im Jahr 2018, in einem Fall sogar erst im Jahr 2019.

Bundesland	Aktualität LoD1	Abdeckung	Aktualität LoD2	Abdeckung	Verfügbarkeit
Baden-Württemberg	2015	100%	2015	100%	AdV-GR
Bayern	2016	100%	2012-2016	75%	AdV-GR
Berlin	2015	100%	-	0%	Frei
Brandenburg	1997-2014	100%	2013-2016	74%	AdV-GR
Bremen	2012,2015	100%	2019	0%	50,- € ¹
Hamburg	2015	100%	2016	0%	Frei
Hessen	2011	100%	2018	70%	AdV-GR
Mecklenburg-Vorpommern	2016	100%	2014	95%	AdV-GR
Niedersachsen	2011-2013	100%	-	0%	KOVerm ²
Nordrhein-Westfalen	2016	100%	2016	100%	Frei
Rheinland-Pfalz	2013	100%	2013-2016	100%	AdV-GR
Saarland	2015	100%	2015	100%	AdV-GR
Sachsen	2009-2014	100%	2009-2014	100%	650 € ³
Sachsen-Anhalt	2016	100%	2016	100%	Frei mit Lizenz ⁴
Schleswig-Holstein	2016	100%	2013-2016	12%	AdV-GR
Thüringen	2016	100%	2015-2016	65%	Frei ⁵

¹ E-Mail Auskunft GeoDatenService Bremen vom 01.06.2017

² E-Mail-Auskunft Landesamt für Geoinformation und Landesvermessung Niedersachsen (LGLN) vom 22.12.2017

³ E-Mail-Auskunft Staatsbetrieb Geobasisinformation und Vermessung Sachsen vom 07.06.2017

⁴ E-Mail-Auskunft Landesamt für Vermessung und Geoinformation Sachsen-Anhalt vom 08.06.2017

⁵ E-Mail-Auskunft Landesamt für Vermessung und Geoinformation Thüringen vom 09.06.2017

Tabelle 3.6: Verfügbarkeiten der 3D-Gebäudedaten in den einzelnen Bundesländern, Stand Dezember 2017

Format Das Format CityGML ist ebenso wie GML ein Abkömmling von XML. Es zeichnet sich dadurch aus, nicht nur zwei-, sondern auch dreidimensionale Objekte abzuspeichern. Dabei werden

^{3,9}E-Mail-Auskunft Landesamt für Vermessung und Geobasisinformation Rheinland-Pfalz vom 08.06.2017

3D-Geometrien, 3D-Topologien, Semantik und das Erscheinungsbild repräsentiert. Es ist definiert als GML3-Anwendungsschema, also als GIS, das in der Lage ist, 3D-Modelle für Stadt und Land zu erfassen und den Austausch zu gewährleisten. Die Themenfelder, die abgedeckt werden, sind vielseitig: Gebäude, Vegetation, Gewässer, Geländereiefs, Verkehr usw. Neben den bereits erwähnten Detailstufen existieren noch die Detailstufe LoD0 mit 2,5-dimensionalen Daten, das LoD3 mit hochauflösenden Abbildungen der äußeren Erscheinungsform, sowie das LoD4, mit dem es möglich ist, virtuell begehbare Gebäudemodelle bereitzustellen und auszutauschen. (Bill 2016, S. 254 ff.)

Das verwendete Koordinatensystem ist auch hier EPSG:25832.

3.3. Bodenrichtwerte

Bodenrichtwerte werden mittlerweile zentral von der AdV-Projektgruppe VBORIS verwaltet. Sie hat es sich zur Aufgabe gemacht, die in den einzelnen Bundesländern vorhandenen Daten zu den Bodenrichtwerten in einem zentralen Gemeinschaftsportal zugänglich zu machen (Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) 2013, S. 1). Die Bodenrichtwerte werden von Gutachterausschüssen festgelegt. Diese sind „neutrale und unabhängige Kollegialgremien und Behörden im Sinne des Verwaltungsverfahrensgesetzes“ (ebd., S. 3). Sie bestehen aus dem vorsitzenden Mitglied, dem Ausschuss als Kollegialorgan und einer Geschäftsstelle. Dabei werden die Zuständigkeitsbereiche, die meist den Landkreisen und kreisfreien Städten entsprechen, von der jeweiligen Landesregierung bestimmt. Die ehrenamtlichen Mitglieder werden von der öffentlichen Verwaltung einberufen und stammen aus fachnahen Kreisen. Die Grundlage für die Bestimmung von Bodenrichtwerten liefern „das Baugesetzbuch (BauGB), die Immobilienwertermittlungsverordnung (ImmoWertV) sowie die Gutachterausschussverordnungen und weitergehende Regelwerke der Länder in der jeweils geltenden Fassung“ (ebd.).

Nach Angabe von Arbeitsgemeinschaft der Vorsitzenden der Gutachterausschüsse für Grundstückswerte in Nordrhein-Westfalen (AGVGA.NRW) (2017) ist ein Bodenrichtwert ein vorwiegend aus Kaufpreisen ermittelter durchschnittlicher Bodenwert je Quadratmeter für ein Gebiet mit im wesentlichen gleichen wertbestimmenden Merkmalen wie z.B. Entwicklungszustand, Art und Maß der baulichen Nutzung sowie Zuschnitt; er ist bezogen auf ein baureifes Grundstück, dessen Eigenschaften für das Gebiet typisch sind (so genanntes Richtwertgrundstück). Bodenrichtwerte beziehen sich grundsätzlich auf unbebaute Grundstücke. Der Bodenrichtwert ist kein Verkehrswert. Abweichungen des zu bewertenden Grundstücks in den wertbestimmenden Eigenschaften wie z.B. Art und Maß der baulichen Nutzung, Bodenbeschaffenheit, Erschließungszustand und Grundstücksgestaltung bewirken Abweichungen seines Verkehrswertes vom Bodenrichtwert. Die Gutachterausschüsse haben bis zum 15. Februar jedes Jahres Bodenrichtwerte, bezogen auf den 01. Januar des laufenden Jahres,

zu ermitteln. Sie sollen bis zum 31. März veröffentlicht werden. Aus allen verfügbaren Bodenrichtwerten wird anschließend ein landeseinheitlicher Datensatz gebildet. Neben der Bereitstellung der aktuellen Bodenrichtwerte werden auch die Daten der vorangegangenen Jahre ab dem Jahr 2011 angeboten. (Arbeitsgemeinschaft der Vorsitzenden der Gutachterausschüsse für Grundstückswerte in Nordrhein-Westfalen (AGVGA.NRW) 2017)

Verfügbarkeit Die Verfügbarkeiten der Bodenrichtwerte in den einzelnen Bundesländern sind unterschiedlich. In Bremen, Bayern, Thüringen und Niedersachsen sind die Kartendaten mit den Bodenrichtwerten kostenpflichtig. In vielen anderen Bundesländern kann über das jeweilige Geoportal die Karte zwar angezeigt werden, aber nicht ohne weiteres heruntergeladen werden. Die Länder Mecklenburg-Vorpommern, Rheinland-Pfalz und Saarland bieten die Möglichkeit an, den angezeigten Kartenausschnitt auf dem eigenen Rechner zu speichern. Einzig das Land Nordrhein-Westfalen bietet den direkten Download einer ESRI-Shape-Datei an. In Tabelle 3.7 sind die Verfügbarkeiten der Bodenrichtwerte aufgelistet. Die Informationen stammen aus den Geoportalen der Länder.

Bundesland	Bodenrichtwerte
Baden-Württemberg	anzeigbar
Bayern	kostenpflichtig
Berlin	anzeigbar
Brandenburg	anzeigbar
Bremen	kostenpflichtig
Hamburg	anzeigbar
Hessen	anzeigbar
Mecklenburg-Vorpommern	Kartenausschnitt
Niedersachsen	kostenpflichtig
Nordrhein-Westfalen	frei
Rheinland-Pfalz	Kartenausschnitt
Saarland	Kartenausschnitt
Sachsen	anzeigbar
Sachsen-Anhalt	anzeigbar
Schleswig-Holstein	anzeigbar
Thüringen	kostenpflichtig

Tabelle 3.7: Verfügbarkeiten der Bodenrichtwerte in den einzelnen Bundesländern, Stand August 2017

Format Das Shape-Format der Firma Esri, die Entwicklerin der Programme ArcGIS und ArcView ist, ist als Quasi-Standard weit verbreitet (Bill 2016, S. 640). Es hat somit die Qualität eines universellen Austauschformates für Vektordaten. Allerdings hat es die Einschränkung, dass z. B. Namen von

Feldern auf 10 Zeichen begrenzt sind. Es sind mindestens drei Dateien nötig, dabei werden in einer .shp-Datei die Geometriedaten gespeichert, in einer .dbf-Datei werden Sachdaten ohne Geometrien gespeichert; eine .shx-Datei dient zur Indizierung für den schnelleren Zugriff auf die vorher genannten Daten und deren Verknüpfung (Bill 2016, S. 640). Die Daten liegen ebenfalls im Koordinatensystem EPSG:25832 vor. Die Daten sind für die Jahre 2011 bis 2017 verfügbar. Um einen engen zeitlichen Bezug zu den Zensus-Daten von 2011 zu erhalten werden die Daten des selben Jahres verrechnet. Der Bodenrichtwert ist in €/m² angegeben.

3.4. Zensus-Daten

Die Zensus-Daten mit dem Ergebnis der Volkszählung von 2011 liegen für die gesamte Bundesrepublik Deutschland vor. Allerdings sind diese maskiert: Gitterzellen mit einzelnen Personen werden wie Gitterzellen ohne Personen klassifiziert und Gitterzellen mit nur zwei Personen wie Gitterzellen mit drei Personen (Statistisches Bundesamt 2016, S. 3). Gitterzellen, deren Einwohnerzahl geheim gehalten werden oder in denen kein Einwohner zu verzeichnen ist, erhalten den Wert von –1 für die Einwohnerzahl. Die Zensus-Daten sind nicht konsistent, nicht nur aufgrund von Messfehlern. Das Grundgesetz der Bundesrepublik Deutschland erlaubt keine allzu scharfe Erfassung. In einem freien Land könne die Einwohnerzahl nie exakt ermittelt werden ^{3.10}.

Verfügbarkeit Die Zensus-Daten liegen für die gesamte Bundesrepublik Deutschland frei verfügbar vor und können von der Website www.zensus2011.de des Statistischen Bundesamtes kostenlos heruntergeladen werden.

Format Die Zensus-Daten liegen im Format CSV vor. CSV steht für *Comma Separated Value* (Bill 2016, S. 640), deutsch: Komma-getrennte Werte, und ist dementsprechend einfach aufgebaut: es kann mit einem beliebigen Text-Editor geladen und geändert werden. Im Fall der Zensus-Datensätze existieren vier Felder: der Name der 100 mal 100 m großen, d. h. ein Hektar großen Rasterzelle, die Koordinaten in x- und y-Ausrichtung des Mittelpunktes der Rasterzelle und die Anzahl der in der Rasterzelle lebenden Einwohner. Die Einwohnerdichten, die sich daraus ableiten lassen, können daher in Einwohner pro Hektar angegeben werden und tragen die Einheit E/ha. Das verwendete Koordinatensystem ist die *Lambert-Azimuthal-Equal Area Projection*, in der die Koordinaten der Mittelpunkte einer Rasterzelle ausgewiesen werden (Statistisches Bundesamt 2016, S. 2). Die Kurzform lautet ETRS89-LAEA Europe oder EPSG:3035.

^{3.10}Innenstaatssekretär Klaus Vitt in Hamburger Abendblatt vom 27.10.2017 <http://www.abendblatt.de/hamburg/article212340783/Hamburg-wehrt-sich-vor-Gericht-gegen-Zensus-Einbussen.html>, besucht am 01.11.2017

3.5. Sonstige Daten

Verwaltungsgrenzen Die Daten für die Verwaltungsgrenzen in Nordrhein-Westfalen liegen ebenfalls im ESRI-Shape-Format vor. Hier gibt es die Objektarten für die Landesgrenze, die Regierungsbezirke, die Kreise sowie die kreisgebundenen und kreisfreien Gemeinden. Die Verwaltungsgrenzen der Gemeinden werden dafür genutzt, die Zensus-Daten für eine Gemeinde aus dem gesamten Datensatz für die Bundesrepublik Deutschland herauszuschneiden, siehe Kap. 4 *Datenaufbereitung und -analyse*. Das Koordinatensystem ist hier EPSG:25832.

4. Datenaufbereitung und -analyse

Die Schritte zur Aufbereitung der Daten und die abschließende Analyse werden im Folgenden beschrieben. Die Datenaufbereitung besteht zum einen aus einer Reduktion der Daten und zum anderen aus einer Synthese. Die Datenaufbereitung wird mit Hilfe der Software FME durchgeführt. Zur Vorbereitung zählt die Auswahl der Geoobjekte über ein Rechteck (vgl. Abschnitt 2.3), das die Anzahl der einzulesenden Objekte vor der eigentlichen Berechnung durch FME erheblich reduziert, vor allem dadurch, dass die Geodaten nach Kreisen paketierte sind und bei kleineren Gemeinden sonst die Daten für den gesamten Kreis eingelesen werden müssten. Eine Ausnahme bilden die vorbereitenden Schritte der Datenreduktion der Zensus-Daten, siehe Abschnitt 4.1 *Datenreduktion*. Alle sonstigen FME-Workspaces werden zu einem gesamten Workspace zusammengeschaltet, wobei jeder Workspace nun als Transformer dargestellt wird, siehe Abbildung 4.2. Die Funktion der einzelnen Transformer respektive Workspaces werden im Abschnitt 4.2 erläutert. Die folgenden Abbildungen sind Bildschirmausschnitte, die bei der Berechnung der Daten für die Gemeinde Alpen entstanden sind. Die Ergebnisse der Berechnungen für die Gemeinde Alpen und die zwei weiteren Gemeinden sind in Kapitel 5 aufgeführt.

Für die Organisation und Unterscheidbarkeit wurde folgende Notation eingeführt: Operationen, die an Daten für die Gebäude durchgeführt werden beginnen mit einem *G*, siehe Abschnitt 4.2.2. Analog dazu sind solche, die sich mit der Berechnung von Flurstücken befassen, mit einem *F* gekennzeichnet, siehe Abschnitt 4.2.1. Vorbereitende Operationen sind mit einem *Z* versehen. Attribute, die mit der FME erzeugt und berechnet werden, erhalten zur Kennzeichnung das Suffix *_fme*, z. B. *height_fme*. Temporäre Attribute, die nur einen Zwischenschritt darstellen, erhalten das Präfix *_*, z. B. *_zmax*.

Die anschließende Analyse der Daten erfolgt in einem weiteren Workspace *C1*. Die Trennung von Datensynthese und -analyse hat den Hintergrund, dass die Software FME sehr rechenintensiv ist und dazu viel Arbeitsspeicher benötigt. Außerdem erlaubt diese Trennung die mehrmalige Durchführung der Clusteranalyse, ohne die gesamten Daten von Grund auf neu zu berechnen. Die Ergebnisse werden anschließend mit Hilfe des Geoinformationssystems QGIS und der Statistik-Software R visualisiert.

Nordrhein-Westfalen mit einem Umfang von 396 Datensätzen, hier als `dvg1gem_nw` bezeichnet. Mit einem Transformer des Typs *AttributeFilter* wird zunächst die Gemeindegrenze (Stadtgrenze) extrahiert. In einem weiteren Schritt wird um die Stadtgrenze eine 200 m breite Pufferzone gelegt. Dieser Vorgang geschieht mit einem Transformer des Typs *Bufferer*. Dies hat den Sinn, dass Rasterpunkte aus den Zensus-Daten, die knapp vor der Stadtgrenze liegen, mitgenommen werden. Diese sind nämlich für die Berechnung der Einwohnerdichten in den Flurstücken (Schritt *F2*), die an der Stadtgrenze liegen, relevant. Da die beiden Datenquellen in unterschiedlichen Koordinatensystemen vorliegen und es für die auszuführende geometrische Verschneidung notwendig ist, dass die Geoobjekte auf dem gleichen Koordinatensystem basieren, stellt sich die Frage, welche Datenquelle vor der Verschneidung reprojiziert wird. Aufgrund der Anzahl der Objekte in den beiden Datenquellen ist es vom Berechnungsaufwand sinnvoller, das Geoobjekt, das die Stadtgrenze darstellt, in das Koordinatensystem der Zensus-Daten zu überführen, als andersherum die rund 3,5 Millionen Datensätze aus dem Zensus von Nordrhein-Westfalen. Die Überführung in das andere Koordinatensystem geschieht mit Hilfe des Transformers *Reprojector*.

Mit einem Transformer des Typs *Clipper* werden dann die Zensus-Daten, die innerhalb der Stadt oder Gemeinde und der Pufferzone liegen, ausgeschnitten. Die so erhaltene reduzierte Anzahl an Zensus-Rasterpunkten wird in die Ausgabedatei *Zensus* zur Weiterverarbeitung im FME-spezifischen FFS-Format geschrieben. Zusätzlich wird mit dem Transformer *BoundsExtractor* die maximale Ausdehnung in x- und y-Richtung des Puffers um die Gemeinde ausgelesen. Die so erhaltenen Minimum- und Maximum-Werte werden für die Reduktion der übrigen Daten auf das Zielgebiet genutzt, siehe nächster Abschnitt.

Z1 - Reduktion der Daten auf Zielgebiet Die bei der Reduktion der Zensus-Daten ebenfalls extrahierten Ausdehnungswerte werden dazu genutzt, um alle weiteren Daten auszuschneiden. Dies geschieht mit einem FME-Workspace, der alle weiteren Daten einliest. Den Readers wird dabei der Parameter für einen sogenannten *Search Envelope*, deutsch: Such-Umschlag, mitgegeben. Dieser benötigt die Angaben zur minimalen und maximalen Ausdehnung jeweils in x- und y-Richtung, sowie des zu verwendenden Koordinatensystems. Die Koordinaten werden dazu als für alle Transformer abrufbare Parameter global für den gesamten Workspace gesetzt. Die einzelnen Reader können nun auf diese Werte zugreifen, ohne dass jedem Reader einzeln diese Koordinaten mitgeteilt werden müssten. Hier werden lediglich die Namen der globalen Parameter als Platzhalter gesetzt. Mit Hilfe dieses vorbereitenden Workspaces werden alle Daten (Bestandsdatenauszug, 3D-Gebäudedaten, Bodenrichtwerte), die sich in dem Rechteck, das die Grenze der Stadt oder Gemeinde umschließt, ausgeschnitten, allerdings nicht abgeschnitten, wenn sie über die Grenze dieses Rechtecks reichen. Damit wird sicher gestellt, dass hier keine Manipulation der Daten stattfindet. Auf eine Abbildung dieses Workspaces wird daher bewusst verzichtet, da ohnehin nichts außer dem soeben geschilderten Vorgang des

Ausschneidens passiert. Eine Besonderheit stellen die 3D-Gebäudedaten dar, da hier Objekte ohne Geometrie und somit auch ohne Raumbezug existieren. Es ist bei dem Format, in dem die Daten zur Verfügung gestellt werden, allerdings so, dass in jeder GML-Datei die gesamte Ausdehnung aller beinhalteten Objekte angegeben ist, und somit die Methode des Rechteck-Ausschneidens auch hier funktioniert.

4.2. Datensynthese

Die Datensynthese ist gegliedert in drei Hauptschritte: Synthese der Flurstücksdaten in den Schritten *F0* bis *F4*, Synthese der Gebäudedaten in den Schritten *G0* bis *G2* sowie der abschließenden Vereinigung der beiden Datensätze in Schritt *GF*. Die Synthese von Daten kann über die direkte Verknüpfung von Datensätzen geschehen oder über komplexere Berechnungsmethoden. Die einfache Verknüpfung kann im folgenden Fall angewendet werden. Es liegen verschiedene Datensätze zu den selben realen Objekten vor. So gibt es Informationen zu Gebäuden als CityGML, also dreidimensionale Daten, als auch Daten aus dem Bestandsdatenauszug, die unter anderem Informationen zu der Art der Gebäude enthalten. Die Verknüpfung dieser Datensätze, auch als Verheiraten oder engl. *merging* bezeichnet, führt zu einer größeren Datengrundlage. Die Verknüpfung kann entweder über eine schon vorhandene Identifikationsnummer (ID) erfolgen, was den Berechnungsaufwand durch FME wesentlich erleichtert, oder über einen räumlichen Zusammenhang, was die Berechnung wesentlich aufwändiger macht, da nicht nur ein einzelner Wert pro Geoobjekt verglichen werden muss, sondern die gesamten Geometriedaten verarbeitet werden müssen. Besteht die Wahl zwischen beiden Methoden, ist die Verknüpfung über eine ID vorzuziehen. Eine komplexere Datensynthese stellt die Projektion der Zensus-Daten auf die Flurstücke dar, siehe Abschnitt *F2* - Projizieren der Zensus-Daten auf Flurstücke.

Der gesamte Workspace der FME ist in Abbildung 4.2 dargestellt. Der Inhalt der einzelnen Transformer wird im Folgenden abgebildet und erklärt.

4.2.1. Synthese der Flurstücksdaten

Die Flurstücke werden zunächst exakt auf die Grenze der Gemeinde/Stadt zugeschnitten. Dies geschieht im Schritt *F0*. Danach werden Flurstücke zusammengefasst, die unter der selben Nummer im Grundbuch geführt werden. Dies geschieht in Schritt *F1*. Die Zensus-Daten werden in Schritt *F2* auf die Flurstücke projiziert. Danach werden die Daten in Schritt *F3* mit weiteren Informationen aus dem Bestandsdatenauszug angereichert. Die Projektion der Bodenrichtwerte auf die Flurstücke geschieht in Schritt *F4*.

F0 - Clippen der Flurstücke Die bereits grob ausgeschnittenen Flurstücke werden in diesem Schritt exakt auf die Grenze der Gemeinde zugeschnitten. Dies wird erreicht über die Filterung nach dem Gemeindeschlüssel. Wie in Kapitel 3.1 *Bestandsdatenauszug* bereits erwähnt, besitzen die Flurstücke ein Attribut namens `gemeindezugehoerigkeit`. Diese Attribut enthält den Verweis auf eine weitere Klasse mit dem Namen `AX_Gemeindezugehoerigkeit`, in dem dann folgende untergeordnete Attribute abgelegt sind: `land`, `regierungsbezirk`, `kreis`, `gemeinde`, `gemeindeteil`. Da der global gesetzte Parameter des Workspaces den kompletten Gemeindeschlüssel enthält, muss dieser

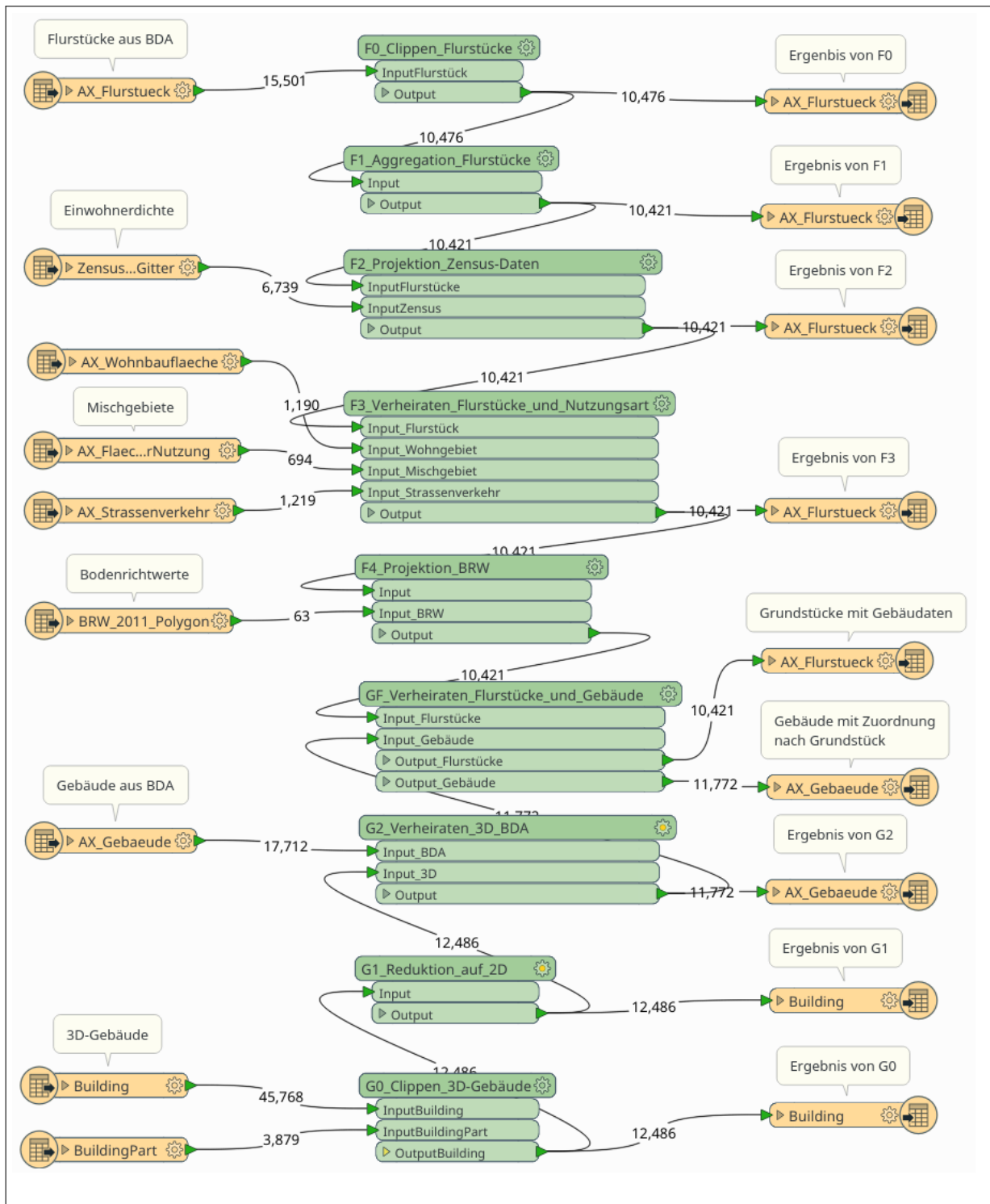


Abbildung 4.2: FME-Workspace zur Datensynthese

nun zerlegt werden. Dies geschieht innerhalb des Transformers `TestFilter`. Es werden Schritt für Schritt die einzelnen Teile dieser Zeichenfolge, engl. *string*, miteinander verglichen. Das Ausschneiden der einzelnen Teile aus dem String, der den Gemeindegeschlüssel enthält, geschieht mit der in FME eingebauten Funktion `substring`. Diese nimmt die folgenden Parameter entgegen: den ursprünglichen String, einen Startwert und eine Längenangabe. Beispielsweise wird dann für den Vergleich mit dem Attribut `land` dann der Startwert auf 0 und die auszulesende Anzahl an Zeichen auf 2 gesetzt, um aus dem Gemeindegeschlüssel den Wert für das Bundesland zu extrahieren, der dann mit dem Inhalt des Attributes `land` verglichen wird. Dies geschieht analog mit allen Attributen zur Gemeindezugehörigkeit. In Abb. 4.3 ist die Umsetzung mit FME dargestellt.

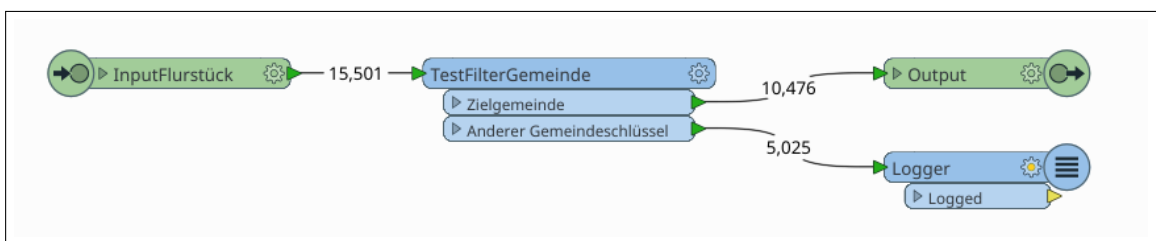


Abbildung 4.3: FME-Workspace *F0* - Clipping der Flurstücke

F1 - Aggregation der Flurstücke Dieser Schritt ermöglicht es, Flurstücke, die unter der selben Nummer im Grundbuchamt geführt werden und die sich aufeinander beziehen, zu einem Grundstück zusammenzufassen. Dies setzt voraus, dass die Listenattribute für die Beziehungen geordnet sind. Dies geschieht mit zwei Transformern des Typs `ListSorter`, die die Attribute `weistAuf` und `zeigtAuf` alphabetisch sortieren. Die eigentliche Zusammenfassung geschieht dann mit einem Transformer des Typs `Aggregator`, der die Flurstücke über die beiden erwähnten Attribute und das Attribut `istGebucht` gruppiert und aggregiert. Der Prozess ist in Abb. 4.4 dargestellt.

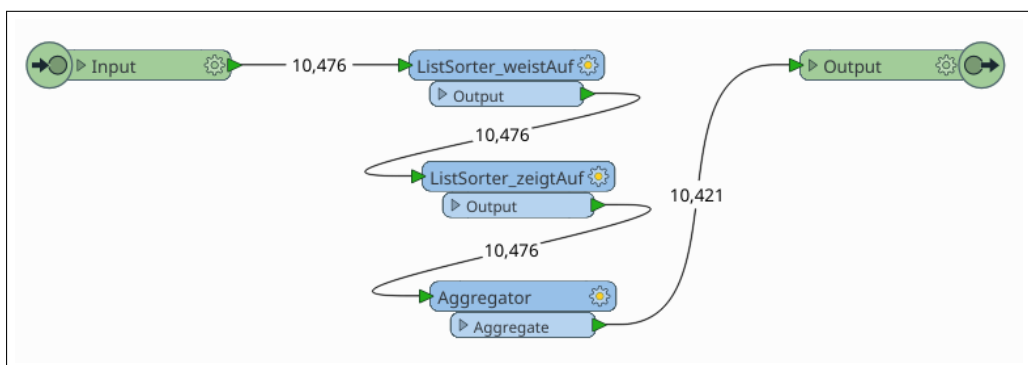


Abbildung 4.4: FME-Workspace *F1* - Aggregation der Flurstücke

F2 - Projizieren der Zensus-Daten auf Flurstücke Um die Einwohnerzahl auf einem Flurstück zu erhalten, müssen die Informationen aus den Zensus-Daten in die Daten zu den Flurstücken einfließen. Die Zensus-Daten liegen wie beschrieben in einem quadratischen Raster vor, bei dem jedes Raster eine Größe von 1 ha aufweist. Die Flurstücke können nicht durch eine einfache geometrische Verschneidung diesen Rasterzellen zugeordnet werden. Außerdem sind starke Sprünge zwischen den einzelnen Rasterzellen zu beobachten. Erschwerend kommt hinzu, dass für dünn besiedelte Gebiete keine Zensus-Daten vorliegen. Der im Zensus-Datensatz vorkommende Wert für die Einwohnerzahl von „-1“ kann entweder bedeuten, dass hier keine Einwohner vorhanden sind, oder dass der Wert maskiert wurde, z. B. aus Gründen der Anonymisierung.

Für die eigentliche Projektion der Einwohnerdichten wird auf die Methode der invers quadratischen Gewichtsfunktion zurückgegriffen, siehe Abschnitt 2.3. Mit ihr können die Einwohnerdichten in den Flurstücken interpoliert werden. Die Zensus-Daten stellen im Grunde keine Einwohnerzahlen dar, sondern Einwohnerdichten, da ja die Einwohnerzahl pro Hektar angegeben ist. Der Workspace ist in Abbildung 4.5 zu sehen. Zunächst werden die Daten zu den Flurstücken aus dem vorangegangenen Abschnitt eingelesen und die Geometriedaten mit dem Transformer *GeometryExtractor* in einem temporären Attribut gesichert und zwischengelagert. Dies ist nötig, da im nächsten Schritt die Geometrie des Polygons, also die Flurstücksgrenze, ersetzt wird durch den Mittelpunkt des Flurstücks. Dies geschieht über einen Transformer des Typs *CenterPointReplacer*. Die Geometrie des Flurstücks besteht jetzt nicht mehr in einem Polygon, sondern nur noch in einem Punkt. Für diesen Punkt werden mit Hilfe des Transformers *NeighborFinder* die nächsten drei Rasterpunkte aus den Zensus-Daten gesucht. Die Zensus-Daten sind dafür vorher mit einem *Reprojector* in das gleiche Koordinatensystem gesetzt worden. Ein Transformer des Typs *AttributeManager* demaskiert die Zensus-Daten. Dabei werden folgende Wenn-Dann-Bedingungen entsprechend den Angaben zu den Zensus-Daten (vgl. Kap. 3 *Datengrundlage*) gesetzt: Wenn die Einwohnerdichte kleiner als 0 ist, setze sie gleich 1; wenn die Einwohnerdichte 0 ist, setze sie gleich 1; wenn die Einwohnerdichte 3 ist, setze sie gleich 2.

Im nächsten Schritt werden die Flurstücke und die Zensus-Daten verknüpft. Der *NeighborFinder* sucht nicht nur die nächsten drei Zensus-Rasterpunkte `_candidates{}`, sondern speichert auch deren Werte bezüglich der Einwohnerzahl `_candidates{ }.Einwohner` und ihre jeweilige Entfernung zum Mittelpunkt des Flurstückes `_candidates{ }.distance` in einem temporären Attribut ab. Die beiden Werte sind für die anschließende invers quadratische Gewichtsfunktion nötig. Die Suche nach den nächsten benachbarten Rasterpunkten kann hier auf 200 m begrenzt werden, da sich aufgrund des Rasters von 100 mal 100 m auf jeden Fall in dieser Entfernung drei Rasterpunkte befinden. Nachdem dies geschehen ist, wird die ursprüngliche Geometrie des Flurstücks mit dem Transformer *GeometryReplacer* wiederhergestellt. Mit einem Transformer des Typs *AttributeManager* wird die Berechnung der Einwohnerdichten mit Hilfe der invers quadratischen Gewichtsfunktion berechnet. Dazu wird ein neues Attribut mit dem Namen `ewdichte_fme` erstellt und nach der Methode der

invers quadratischen Gewichtungsfunktion mit Daten gefüllt. Die Befehlszeile zur Berechnung, d. h. die Umsetzung der invers quadratischen Gewichtungsfunktion lautet in der FME-eigenen Skript-Sprache:

```
@div(
  @sum(
    @Value(_candidates{0}.Einwohner)/@Value(_candidates{0}.distance)**2,
    @Value(_candidates{1}.Einwohner)/@Value(_candidates{1}.distance)**2,
    @Value(_candidates{2}.Einwohner)/@Value(_candidates{2}.distance)**2),
  @sum(
    1.0/@Value(_candidates{0}.distance)**2,
    1.0/@Value(_candidates{1}.distance)**2,
    1.0/@Value(_candidates{2}.distance)**2))
```

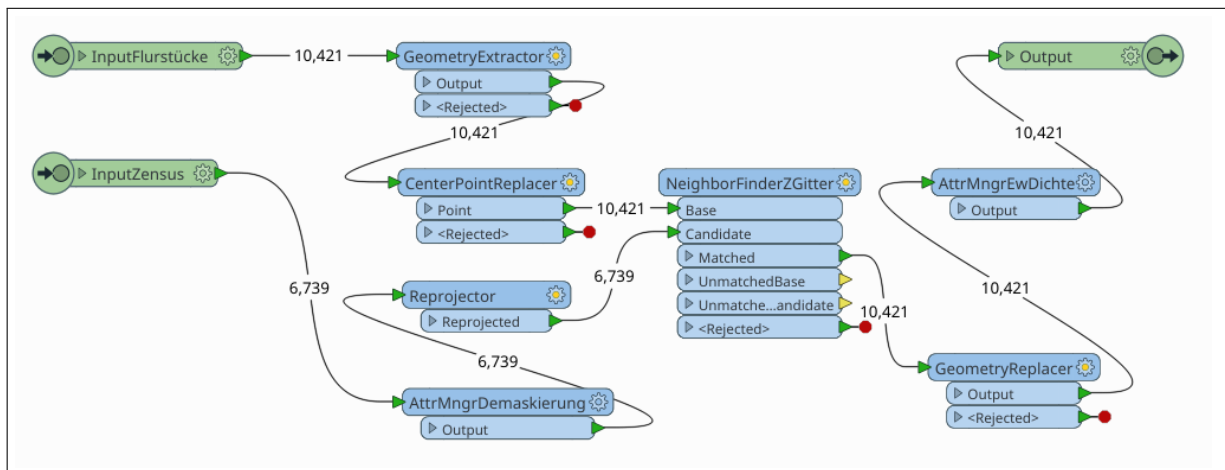


Abbildung 4.5: FME-Workspace F2 - Projizieren der Zensus-Daten auf Flurstücke

F3 - Verheirateten Flurstücke und Nutzungsart Zu den Flurstücken liegen verschiedene Datenquellen vor. So gibt es allein in den Bestandsdaten drei verschiedene Datensätze, die relevant sind. Dazu zählen die eigentlichen Daten zu den Flurstücken (AX_Flurstueck), die Daten zu den Flurstücken, die die Information enthalten, ob ein Flurstück eine Wohnbaufläche ist (AX_Wohnbauflaeche), ob es eine Fläche mit gemischter Nutzung (AX_FlaecheGemischterNutzung) ist, oder eine Verkehrsfläche (AX_Strassenverkehr). Alle Datenquellen liegen im gleichen Koordinatensystem vor, was eine Reprojektion überflüssig macht. Die Daten unterliegen allesamt den AAA-Richtlinien, so dass sie überschneidungsfrei und lückenlos sind. Dies ermöglicht die Verschneidung der Geodaten über die Transformer des Typs *SpatialRelator*. Dieser verknüpft Daten aufgrund einer geometrischen Beziehung. Da die Daten nicht über eine gemeinsame Identifikationsnummer verknüpft sind, ist dieser Schritt

notwendig. Damit die Identifikationsnummern, die sich in allen Datensätzen in einem Attribut namens `gml_id` befinden, sich nicht gegenseitig überschreiben können, damit später auf sie zurückgegriffen werden kann, werden mit den Transformern des Typs *AttributeManager* diese umbenannt. Die Flurstücke enthalten danach ihre ID im Attribut `flur_gml_id`, die Wohnbauflächen in `wohn_gml_id`, die Mischgebiete in `mix_gml_id` und die Verkehrsflächen im Attribut `strv_gml_id`.

Da die Polygone der Wohnbauflächen, der Flächen mit gemischter Nutzung und Verkehrsflächen auf den Polygonen der Flurstücke basieren können, oder Zusammenfassungen von Flurstücken sein können, oder ganz andere Grenzen aufweisen können, müssen auch diese verschiedenen topologischen Beziehungen berücksichtigt werden. Die drei Transformer des Typs *SpatialRelator* prüfen nun, in welcher Beziehung sich ein Polygon aus den Daten der Flurstücke zu den Polygonen der Wohnbauflächen, der Flächen gemischter Nutzung oder der Verkehrsflächen befindet. Dies geschieht für Wohnbauflächen und Mischflächen mit der Anweisung `Requestor intersects Supplier`. Hier wird also die geometrische Beziehung *intersect* angewendet, vgl. Kap. 2.3 *Geoinformationssysteme*. Für die Verkehrsflächen ist es wichtig, dass diese nicht durch eine bloße Überschneidung, wie im vorherigen Schritt, sondern durch eine Umschließung identifiziert werden. Daher wird hier dem *SpatialRelator* die Anweisung gegeben, zu prüfen, ob sich das Flurstück in der Verkehrsfläche befindet (`Requestor is Within Supplier`) oder umgekehrt (`Requestor Contains Supplier`). Ansonsten würden viele Flurstücke, die eine kleine Überschneidung mit einer Verkehrsfläche haben, auch in Gänze einer solchen zugeteilt. Der FME-Workspace zum aktuellen Schritt ist in Abbildung 4.6 dargestellt.

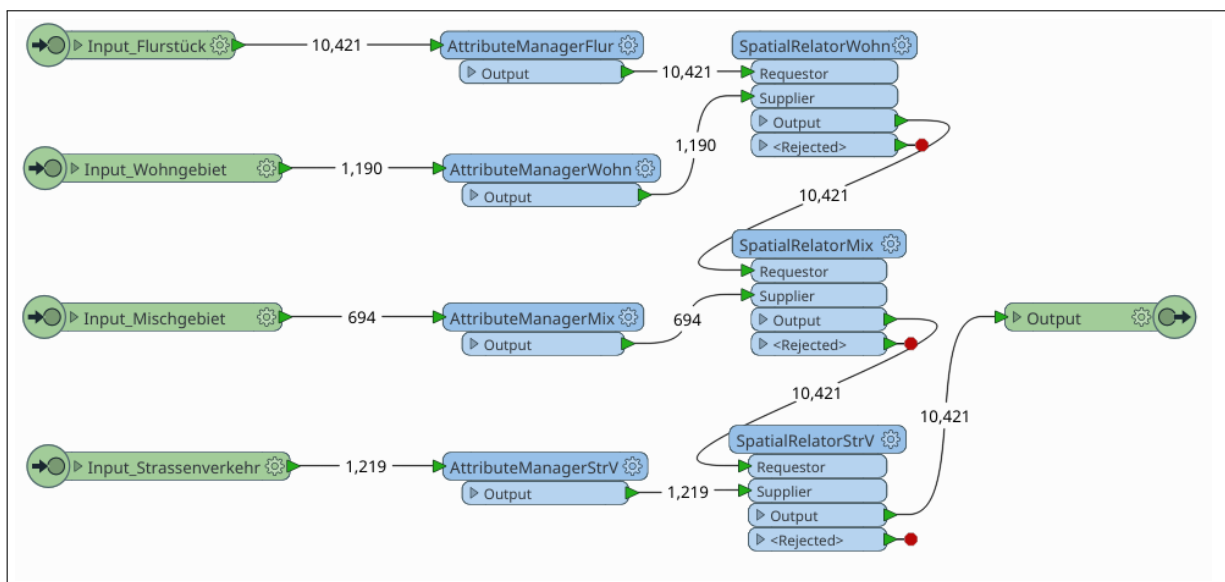


Abbildung 4.6: FME-Workspace F3 - Verheiraten Flurstücke und Nutzungsart

F4 - Projizieren der Bodenrichtwerte auf Flurstücke Die Projektion der Bodenrichtwerte (BRWe) erfolgt analog zur Projektion der Zensus-Daten mit Hilfe der invers quadratischen Gewichtsfunktion. Da die BRWe nicht in einem regelmäßigen Raster vorliegen, sondern unterschiedlich große und geartete Polygone sind, kann hier nicht die Suche nach den nächsten Nachbarn auf eine maximale Distanz begrenzt werden. Da die BRWe als Kommazahlen in einer Zeichenfolge angeben sind, müssen diese zuerst in einen anderen Datentyp umgewandelt werden, damit FME die Werte als Zahlen erkennt. Dies geschieht mit einem Transformer des Typs *AttributeManager*, der mit Hilfe einer Konvertierungsfunktion diesen Schritt durchführt. Der Befehl in der FME-Sprache lautet dafür: `@double(@Value(BRW))`. Alle anderen Vorgänge laufen so ab wie in Schritt F2. Der Workspace ist in Abb. 4.7 zu sehen. Damit ist die Berechnung für die Flurstücke abgeschlossen.

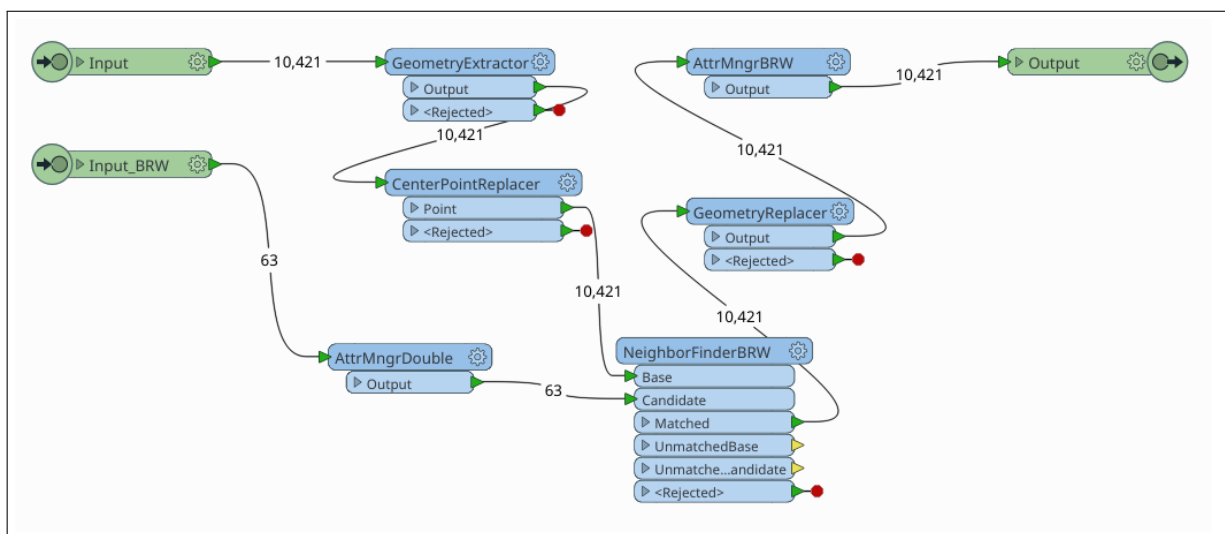


Abbildung 4.7: FME-Workspace F4 - Projizieren der Bodenrichtwerte auf Flurstücke

4.2.2. Synthese der Gebäudedaten

Die Daten der Gebäude stammen zum einen auch aus dem Bestandsdatenauszug, zum anderen sind es 3D-Gebäudedaten. Letztere werden zunächst für die Stadt/Gemeinde ausgewählt, dies geschieht in Schritt G0. Dabei wird gleichzeitig die Datenstruktur verändert. Aus den 3D-Daten werden im nächsten Schritt G1 geometrische Daten und Kennwerte extrahiert, so dass die eigentliche dreidimensionale Struktur verworfen werden kann. Als letzter Schritt G2 werden die Daten mit dem Bestandsdatenauszug verknüpft, was dazu führt, dass nur amtlich existente Gebäude zur weiteren Berechnung herangezogen werden.

G0 - Clippen der 3D-Gebäude und Reduktion der XML-Ebenen XML-Dateien, also auch alle Abkömmlinge, wie CityGML und Inspire GML, beinhalten Hierarchien. Es gibt Eltern-Objekte und Kind-Objekte. Da FME zwar alle möglichen Formate einlesen kann, sich aber intern in der Verarbeitung solcher Hierarchieebenen schwer tut, werden einige Datensätze vor der weiteren Verarbeitung auf eine Ebene reduziert. Bei der Datengrundlage zu den Gebäudedaten ist aufgefallen, dass einige Datensätze zu den Gebäuden keine eigene Geometriedaten beinhalten, sondern dass diese in Kind-Objekte ausgelagert sind.

Die CityGML Daten des Landes Nordrhein-Westfalen liegen teilweise in hierarchischer Form vor. So gibt es Feature-Objekte, die für ein Gebäude (**Building**) stehen und solche, die Gebäude-Teile (**BuildingPart**) darstellen. Tritt ein Feature-Objekt des Typs **BuildingPart** auf, besitzt dieses ein Attribut, der auf das Eltern-Objekt verweist. Das jeweilige Elternobjekt ist vom Typ **Building** und besitzt in diesem Fall keine eigenen Geometrie-Daten. Es gibt in jedem CityGML-Datensatz auch noch ein Objekt des Typs **CityModel**, das Meta-Informationen über den gesamten Datensatz enthält. Zugleich ist es Elternobjekt aller Feature des Typs **Building**.

Um die Hierarchie aufzulösen, wurde ein FME-Workspace erstellt, der in der Lage ist, die CityGML-Dateien einzulesen, zu verarbeiten und schließlich ohne Hierarchie weiterzugeben. Die grafische Struktur dieses Workspaces ist in Abbildung 4.8 dargestellt. Die primäre Aufgabe des Zusammenführens wird von einem Transformer des Typs *FeatureMerger* durchgeführt. Vorher werden allerdings alle Objekte des Typs **Building**, die eine eigene Geometrie besitzen und somit auch keine Elternobjekte darstellen, herausgefiltert und parallel zu den übrigen Objekten weiterverarbeitet. Im gleichen Transformer des Typs *TestFilter* werden dabei auch nur die Gebäude gefiltert, die den gesuchten Gemeindeschlüssel enthalten. Dieser ist in jedem Objekt des Typs **Building** abgelegt. In den Objekten des Typs **BuildingPart** ist dieser allerdings nicht hinterlegt, was durch die Verknüpfung über die ID nicht nötig ist.

Um zu vermeiden, dass die gleichnamigen Variablen `gml_id` überschrieben, bzw. verwechselt werden, wird zunächst das Attribut der Gebäudeteile in `gml_buildingpart_id` umbenannt. Mit den Attributen der Eltern-Objekte wurden die **BuildingPart**-Objekte zunächst in einem ersten Schritt mit Hilfe eines *FeatureMerger* angereichert. Die Verknüpfung zwischen *Supplier* und *Requestor* wurde hier durch die 1:n-Beziehung, die ja in der GML-Struktur vorgegeben ist, über die Attribute `gml_parent_id` (für den *Requestor* **BuildingPart**) sowie `gml_id` (für den *Supplier* **Building**) hergestellt. In einem zweiten Schritt wurden dann die Geometrien aus den **BuildingPart**-Objekten mit dem selben Eltern-Objekt über den Transformator *Aggregator* in ein gemeinsames Objekt geschrieben. Um später auf die aggregierten Gebäudeteile zurückgreifen zu können, werden durch den *Aggregator* deren Identifikationsnummern (`gml_buildingpart_id`) in eine Liste (`buildingparts`) geschrieben. Logischerweise entspricht die Anzahl der verbliebenen Objekte der Anzahl der Objekte

des Typs Building ohne Geometrie, die jetzt durch die Aggregation der BuildingPart-Objekte ausgetauscht worden sind.

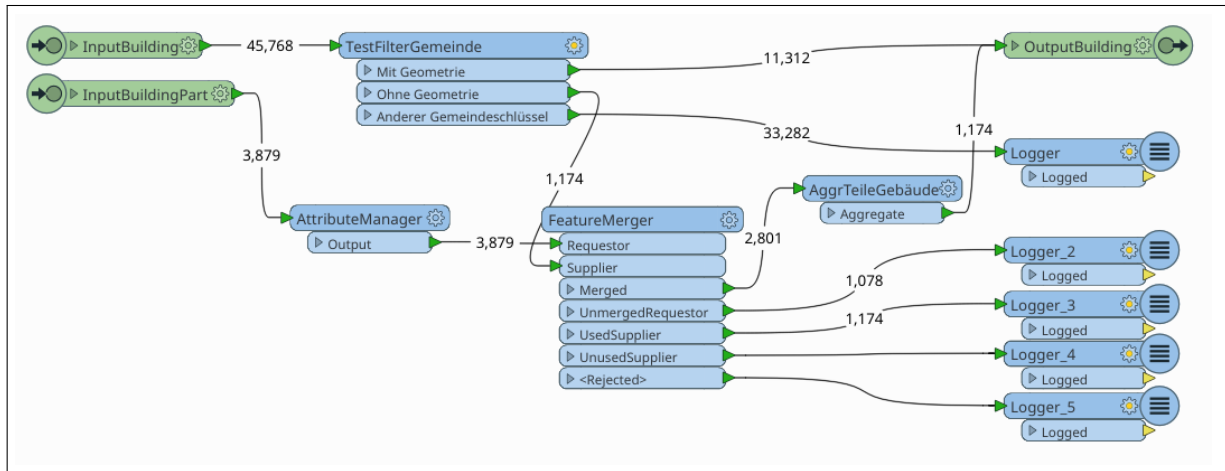


Abbildung 4.8: FME-Workspace G0 - Clippen der 3D-Gebäude und Reduktion der XML-Ebenen

G1 - Reduktion der 3D-Daten auf 2D Die Reduktion auf die zweidimensionale Ebene hat den Vorteil, dass mit einem herkömmlichen Geoinformationssystem, welches nicht 3D-fähig ist, die Daten eingelesen und dargestellt werden können. Zusätzlich verringert sich die Dateigröße. Es werden ausschließlich die Informationen über die Geometrie verringert, die Anzahl der Objekte bleibt gleich. Auch hier wurde zunächst der Datensatz für die Gemeinde Alpen berechnet. Die Anzahl der Objekte entspricht daher der Anzahl der ausgegebenen Objekte im letzten Schritt. Der Workspace in FME ist in Abb. 4.9 dargestellt.

Vor der Reduktion der 3D-Gebäudedaten in die zweidimensionale Ebene werden zunächst Volumen, Außenflächen und Höhe der dreidimensionalen Objekte bestimmt. Für die Berechnung des Volumens wird ein Transformer des Typs *VolumeCalculator* benutzt. Dieser berechnet das Volumen für jedes einzelne Objekt und speichert den Wert in einem neuen Attribut mit dem Namen `volume3d_fme`. Genauso wie für das Volumen, wird auch die Außenfläche jedes 3D-Objekts berechnet. Hierfür wurde ein Transformer des Typs *AreaCalculator* verwendet. Auch hier wird der berechnete Wert in einem neuen Attribut gespeichert. Zur Unterscheidung zur später errechneten Grundfläche, erhält dieses den Name `area3d_fme`. Zur Bestimmung der Höhe der Gebäude werden zunächst mit dem Transformer *BoundsExtractor* die Extremwerte der Ausdehnungen berechnet. Der Transformer speichert u. a. die niedrigste und die höchste Koordinate des Objekts in den temporären Attributen `_zmin` und `_zmax`.

Anschließend erfolgt die Reduktion auf die zweidimensionale Ebene. Dies geschieht mit einem Transformer des Typs *SurfaceFootprintReplacer*. Dieser projiziert das von oben betrachtete Objekt auf die

zweidimensionale Ebene, so dass die Geometrie nur noch aus einer Fläche besteht, deren Flächen-normale nach oben zeigt. Das wichtige an diesem Verfahren ist, dass nicht nur alle Informationen bezüglich der Höhenkoordinate (z-Achse) gelöscht werden, sondern dass die errechnete Fläche nur eine Seite besitzt. Andernfalls würde die weitere Berechnung mit einer doppelt so großen Fläche als der tatsächlichen Grundfläche durchgeführt werden. Mit einem weiteren Transformer des Typs *AreaCalculator* kann nun die Grundfläche des Gebäudes berechnet werden. Das Ergebnis wird in ein neues Attribut `area2d_fme` geschrieben.

Mit einem Transformer des Typs *AttributeManager* werden dann die für diesen Schritt abschließenden Berechnungen durchgeführt. Von der Außenfläche wird die Grundfläche abgezogen, da sie ja streng genommen nicht Teil der Außenfläche ist. Dies geschieht mit dem Befehl `@sub(@Value(area3d_fme), @Value(_area2d_fme))`. Aus der minimalen und maximalen z-Koordinate wird die Höhe (`height_fme`) berechnet: `@sub(@Value(_zmax), @Value(_zmin))`. Die Fläche, die für die spätere Berechnung der Geschossflächenzahl (GFZ) benötigt wird, setzt sich aus der Grundfläche, der Höhe des Gebäudes und der Annahme, dass ein Geschoss eine durchschnittliche Höhe von 3 m besitzt, zusammen. Die Höhe des Gebäudes wird durch drei geteilt und anschließend mit der Grundfläche multipliziert: `@mult(@Value(area2d_fme), @div(@Value(height_fme), 3))`.

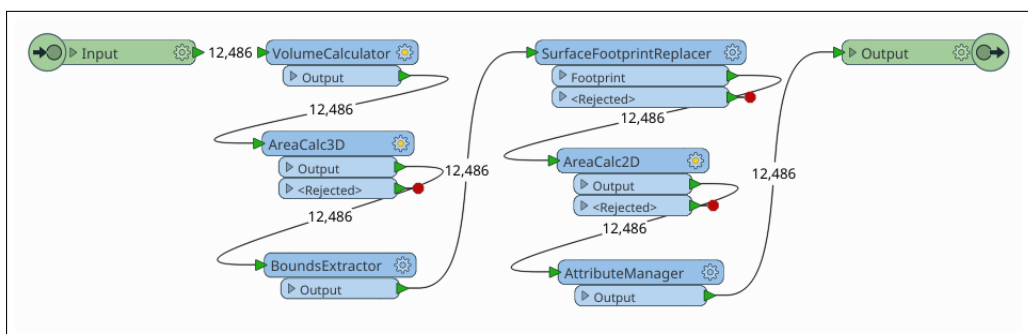


Abbildung 4.9: FME-Workspace G1 - Reduktion der 3D-Daten auf 2D

G2 - Verheiraten von Gebäudedaten Die Gebäudedaten liegen in zwei verschiedenen Quellen mit unterschiedlichen Daten vor. Zum einen werden die schon im letzten Abschnitt auf die zweidimensionale Ebene reduzierten Gebäudedaten eingelesen (*Building*), zum anderen die Gebäudedaten aus dem Bestandsdatenauszug (*AX_Gebaeude*). Über die gemeinsame Identifikationsnummer, intern als `gm1_id` bezeichnet, können die Datensätze miteinander verknüpft werden. Dies geschieht mit einem Transformer des Typs *FeatureMerger*, der angewiesen wird, über die gemeinsame `gm1_id` die Daten zu verbinden. Das Verfahren ist in Abbildung 4.10 dargestellt. Da der Bestandsdatenauszug eine amtliche Datengrundlage darstellt, können mit ihm die Gebäude aus dem CityGML-Datensatz eliminiert werden, die keine amtliche Adresse besitzen. Diese sind meist kleinere Bauwerke, die ohnehin nicht zum Wohnen genutzt werden. Auch werden Gebäude eliminiert, zu denen keine dreidimensionalen Geometriedaten vorliegen. Die verbliebenen Datensätze beinhalten dann die Daten

aus dem Berechnungsschritt im letzten Schritt sowie die Daten aus dem Bestandsdatenauszug. Die übrigen Datensätze werden verworfen.

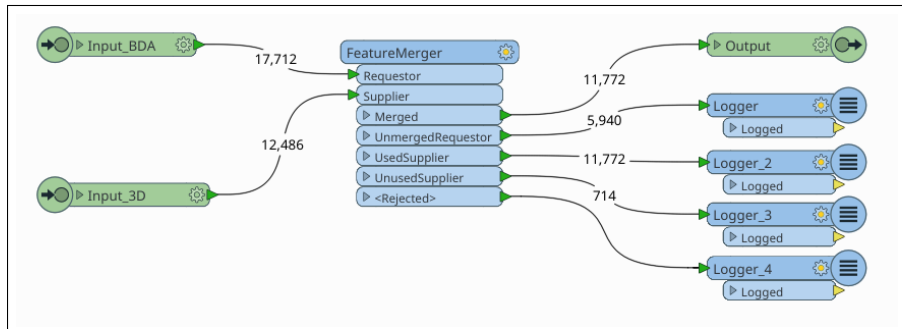


Abbildung 4.10: FME-Workspace G2 - Verheiraten von Gebäudedaten

4.2.3. GF - Verheiraten der Daten - Gebäude und Flurstücke

Die Daten für die Gebäude und Flurstücke werden miteinander verheiratet, um beispielsweise die freie Grundstücksfläche sowie weitere siedlungsspezifische Kennzahlen zu extrahieren. Die Art der Verschneidung ist geometrischer Natur, da es keinen gemeinsamen Wert gibt. Hierzu schreibt die Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) (2008, S. 189): „Die Zuordnung des ‚Gebäudes‘ zum ‚Flurstück‘ kann durch geometrische Verschneidungsoperationen realisiert werden; das explizite Führen von Relationen zwischen den beiden Objektarten unterbleibt.“ Für das weitere Vorgehen ist es also notwendig, aus der geometrischen Relation der beiden Objektarten eine Verknüpfung über ein Schlüsselattribut zu realisieren. Dabei bleiben beide Datensätze erhalten, d. h. es wird in beiden Datensätzen ein neues Attribut erzeugt, in dem vermerkt ist, auf welches Objekt sich das jeweilige Objekt bezieht. Der Workspace ist in Abbildung 4.11 dargestellt.

Im Falle der Gebäude ist dies ein einfaches Vorgehen. Hier wird zunächst im Datensatz der Flurstücke die ursprüngliche Identifikationsnummer `gml_id` mit Hilfe eines *AttributeManager* umbenannt in `flur_gml_id`. Durch einen *SpatialRelator* werden dann die Attribute der Gebäude und Flurstücke über die Bedingung verknüpft, dass Gebäude, die innerhalb eines bestimmten Flurstücks liegen, dessen Attribut `flur_gml_id` zugewiesen bekommen. Da ein Gebäude bzw. dessen Mittelpunkt nur auf einem einzigen Grundstück liegen kann, ist dies ein eindeutiger Zusammenhang.

Anders sieht es im Falle der Anreicherung der Flurstücksdaten mit den Gebäudedaten aus. Hier kann eine beliebige Anzahl von Gebäuden auf einem Grundstück liegen. Außerdem müssen noch die Kennzahlen extrahiert werden, die erst durch die Verknüpfung von Gebäuden und Flurstücken berechnet werden können. Zuerst wird durch einen Transformer des Typs *AreaCalculator* die Fläche des

Flurstücks berechnet. Der Wert wird in einem Attribut mit dem Namen `areaFlur_fme` gespeichert. Anschließend erfolgt die Verknüpfung zwischen den Flurstücken und den Gebäuden. Dies geschieht auch hier mit einem *SpatialRelator*, allerdings mit dem Unterschied, dass die Identifikationsnummern der auf einem Flurstück liegenden Gebäude nicht in einem einfachen Attribut abgelegt werden, sondern in einer Liste. Diese Liste kann beliebig viele Werte aufnehmen, was die Möglichkeit widerspiegelt, dass beliebig viele Gebäude auf einem Flurstück liegen können. Um nun die Kennzahlen zu berechnen ist es notwendig, den Anteil der Summe der Grundfläche und der Geschossflächen der Gebäude an der gesamten Fläche der Flurstücke in Erfahrung zu bringen. Da die Anzahl der Gebäude beliebig sein kann, wird dafür ein sogenannter *CustomTransformer* erstellt, mit dem eine Schleife konzipiert wird. Das Vorgehen ist wie folgt: zuerst wird ein Attribut erstellt, mit dem die Schleifendurchgänge gezählt werden. Dieses Attribut stellt somit einen Iterator dar und wird mit einem *AttributeManager* erstellt. Es wird mit dem Namen `_i` versehen und bekommt als Anfangswert 0 zugewiesen. Zusätzlich werden noch Attribute erstellt, die dazu dienen, die Grundflächen und Geschossflächen der Gebäude zu summieren. Diese werden als `area2d_fme` und `areaGFZ_fme` bezeichnet.

Nun wird die Schleife in Form des *CustomTransformer*, hier als *LoopGebaeude* bezeichnet, betreten. Der Transformer stellt einen eigenen Workspace dar, der in Abbildung 4.12 zu sehen ist. Zuerst wird geprüft, ob die Schleife überhaupt durchlaufen werden soll. Dies geschieht mit einem Transformer des Typs *Tester*. Die Schleife wird dann verlassen bzw. erst gar nicht betreten, wenn der Iterator, der am Ende jedes Schleifendurchlaufs um den Wert 1 erhöht wird, schon den gleichen Wert wie die Anzahl der in der Liste gespeicherten Gebäude entspricht. Wenn der Wert noch unter der Anzahl der Gebäude liegt wird die Schleife durchlaufen. Mit einem Transformer des Typs *ListIndexer* wird dann das Gebäude aus der Liste, dessen Listen-Index dem aktuellen Wert des Iterators entspricht, geladen. Dessen Grundfläche und Geschossfläche wird zu den aktuellen Werten der Attribute `area2d_fme` und `areaGFZ_fme` addiert. Dazu werden neue Attribute erzeugt: `area2d_sum_fme` und `areaGFZ_sum_fme`. So werden bei jedem Durchgang die Grundflächen bzw. Geschossflächen der Gebäude auf einem Flurstück aufsummiert. Gleiches geschieht auch mit den Attributen `volume3d_fme` und `are3d_fme`, wobei die Summen dann in den neuen Attributen `volume3d_sum_fme` und `are3d_sum_fme` gespeichert werden. Anschließend wird der Iterator um den Wert 1 erhöht und die Schleife wird von vorne durchlaufen. Dies geschieht mit dem Transformer *Loop to Input*. Wenn alle Gebäude durch die Schleife gelaufen sind, der Iterator also den gleichen Wert wie die Anzahl der Gebäude in der Liste erreicht hat, wird die Schleife über den Transformer *Output* verlassen.

Wieder im übergeordnete Workspace angelangt, können nun die Grundflächenzahl als Verhältnis der Summe der Grundflächen der Gebäude zur gesamten Fläche des Flurstücks (`@div(@Value(area2d_sum_fme), @Value(area_flurstueck_fme))`) und die Geschossflächenzahl als Verhältnis der Summe der Geschossflächen zur Fläche des Flurstücks (`@div(@Value(areagfz_sum_fme), @Value(area_flurstueck_fme))`) berechnet werden. Auch das Oberflächen-Volumen-Verhältnis wird hier, für das gesamte Flurstück berechnet. Hierzu wird

die Summe aller Flächen durch die Summe aller Volumina geteilt ($\text{@div}(\text{@Value}(\text{area3d_sum_fme}), \text{@Value}(\text{volume3d_sum_fme}))$). Um eine Division durch 0 zu vermeiden, wird der FME über eine Wenn-Dann-Bedingung mitgeteilt, dass bei einer Summe der Oberflächen von 0 ohne weitere Berechnung ebenfalls eine 0 ausgegeben wird. Die Baumassenzahl wird berechnet, indem die Summe der Volumina durch die Fläche des Flurstücks geteilt wird ($\text{@div}(\text{@Value}(\text{volume3d_sum_fme}), \text{@Value}(\text{area_flurstueck_fme}))$). Hier ist die Möglichkeit einer Division durch 0 nicht gegeben, da ein Flurstück immer eine Fläche größer 0 besitzt. Die Flurstücke führen nun die folgenden Kennzahlen als Attribute mit sich: Grundflächenzahl (GRZ) grz_fme , Geschossflächenzahl (GFZ) gfz_fme , Baumassenzahl (BMZ) varatio_fme und Kompaktheit avratio_fme .

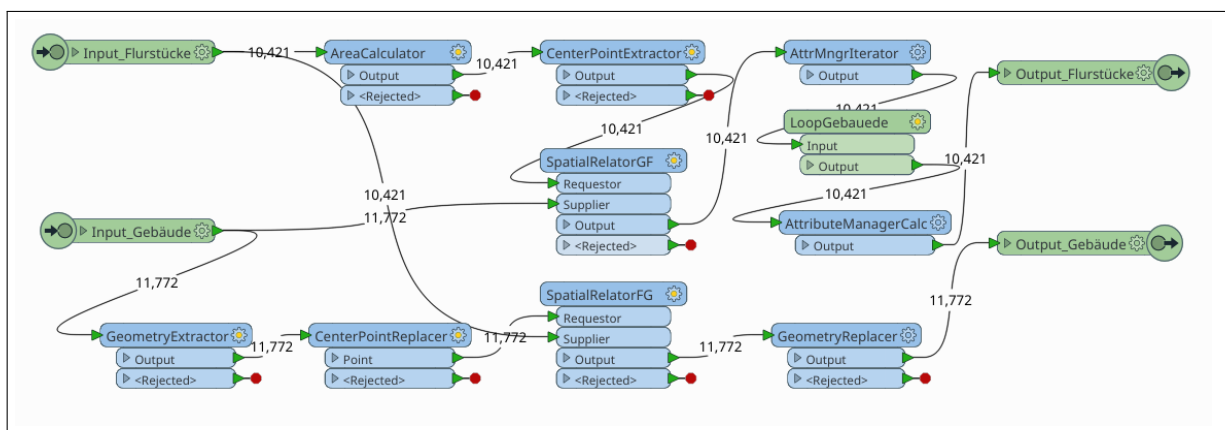


Abbildung 4.11: FME-Workspace GF - Verheiraten der Daten - Gebäude und Flurstücke

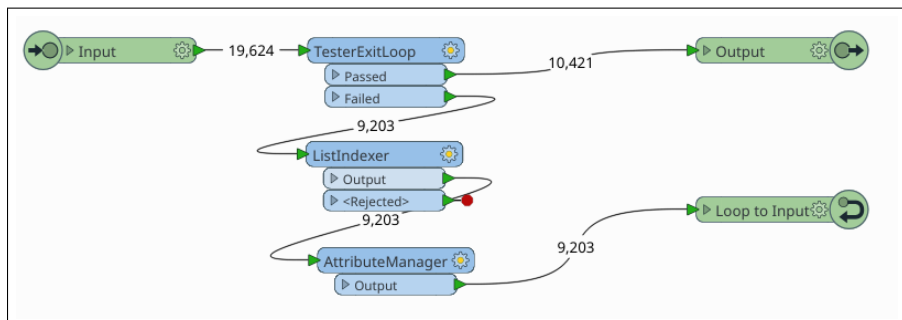


Abbildung 4.12: FME Custom Transformer LoopGebäude

4.3. Datenanalyse

Die Datenanalyse findet zum Teil in der FME statt. Hier wird eine automatisierte Clusteranalyse durchgeführt. Dies geschieht durch die Einbindung der Statistik-Software R mithilfe eines speziellen Transformers. Nach der Clusteranalyse werden die Daten mit den generierten Clustern in die Software R direkt eingelesen, da hier mehr Möglichkeiten der (grafischen) Ausgabe von Ergebnissen existieren. Das detaillierte Vorgehen wird in diesem Kapitel beschrieben.

C1 - Clusteranalyse Für die bereits in Abschnitt 2.4.3 theoretisch beschriebene Clusteranalyse wurde der Workspace *C1 - Clusteranalyse* angelegt. Dieser ist in Abbildung 4.13 dargestellt. Nach Eingang in diesen Transformer laufen die Features zunächst auf einen Transformer des Typs *TestFilter* auf. Dieser prüft, ob die Flurstücke überhaupt bebaut sind. Dann durchlaufen die Features einen weiteren Filterprozess. Hier werden die Features der Flurstücke nach ihrer im Transformer *F3 - Verheiraten Flurstücke* und Nutzungsart zugewiesenen Nutzungsart gefiltert nach Wohngebieten und Mischgebieten. Es hat sich allerdings gezeigt, dass Mischgebiete aufgrund ihrer unterschiedlichen Struktur nicht für eine Clusteranalyse geeignet sind. Daher wird die Clusteranalyse nur für Wohngebiete durchgeführt, deren Anzahl deutlich höher liegt als die Anzahl der Mischgebiete. Zusätzlich prüft der *TestFilter*, ob dem Flurstück auch die Eigenschaft eines Verkehrsweges zugewiesen wurde. Diese Features werden ebenfalls nicht weiter verwendet. Die Features laufen dann zu den Transformern des Typs *RCaller* und gleichzeitig zu den Transformern des Typs *FeatureMerger*. dies hat den Vorteil, dass durch die *RCaller* nur die für die Clusteranalyse notwendigen Attribute laufen müssen, diese aber durch die *FeatureMerger* nach der Berechnung durch R wieder mit den übrigen Daten und Attributen zusammengeführt werden. Außer den Attributen, die direkt für die Clusteranalyse herangezogen werden, wird das Attribut *flur_gml_id* durch den *RCaller* geschleift, damit die *FeatureMerger* über dieses Attribut die Features wieder komplettieren können. Dementsprechend haben die *FeatureMerger* die Einstellung, dass für Requestor und Supplier das Attribut *flur_gml_id* als verknüpfendes Attribut genutzt wird.

Die *RCaller* werden mit den Attributen für die Grundflächenzahl (GRZ) *grz_fme*, für die Geschossflächenzahl *gfl_z*, für das Verhältnis des umbauten Raumes zur Fläche des Flurstücks *varatio_fme*, sowie den Koordinaten des Mittelpunktes des Flurstücks *cpx_fme* und *cpy_fme* beladen. Das Verhältnis der gesamten Oberfläche der Gebäude, abzüglich Grundfläche der Gebäude, zum gesamten Volumen der Gebäude auf dem Flurstück *avratio_fme* wird nicht genutzt, da es sich gezeigt hat, dass die Berechnung der Außenflächen durch die FME fehlerhaft ist, siehe Abschnitt 5.3 *Ergebnisse und Diskussion der Datenaufbereitung*. Zusätzlich wird auch noch das Attribut *flur_gml_id* übergeben, das für die Clusteranalyse nicht herangezogen wird, sondern für die nachgeschalteten *FeatureMerger* nötig ist. Die Clusteranalyse geschieht mit einem R-Script, das in Abb. 4.14 darge-

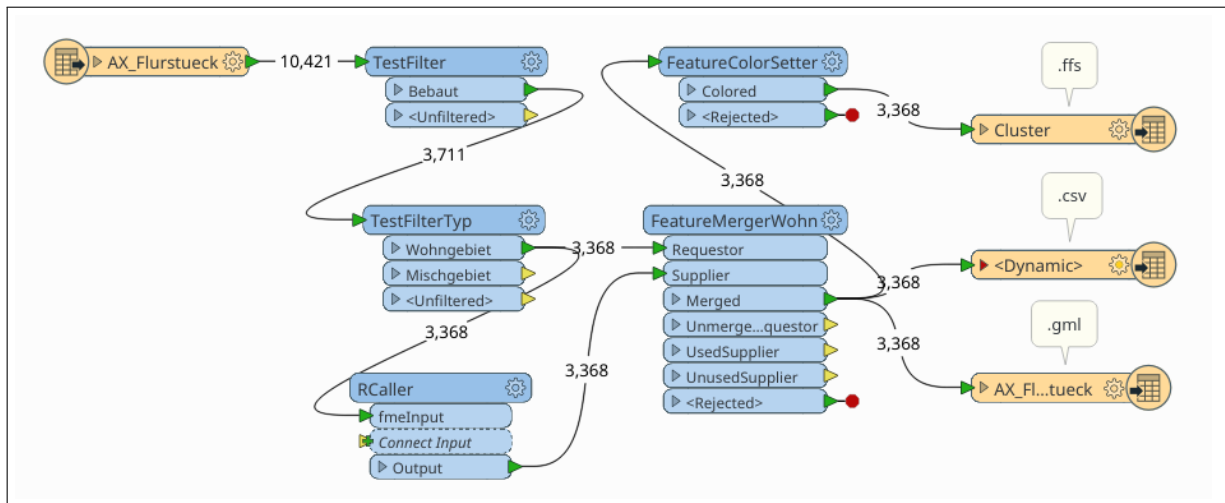


Abbildung 4.13: FME Workspace C1 - Clusteranalyse

stellt ist. Es wird direkt in den `RCaller` eingetragen. In den Zeilen 1 bis 3 werden die benötigten Pakete beziehungsweise Bibliotheken geladen. Die Bibliothek `dplyr` stellt den Befehl `select` aus Zeile 4 zur Verfügung. Die Bibliothek `cluster` beinhaltet den Befehl `daisy` in Zeile 6. Der Befehl `NbClust` in Zeile 7 ist, wie der Name schon vermuten lässt, im Paket `NbClust` enthalten. Mit dem Befehl `select` aus Zeile 4 und 5 wird ein neues Objekt mit dem Namen `wg` erstellt, in das die zur Clusteranalyse benötigten Variablen aus dem Input des Transformers `fmeInput` geladen werden. In Zeile 6 wird ein weiteres Objekt erzeugt, das den Namen `dist.euclid` trägt. Dieser Name wurde deswegen gewählt, da es sich um die Distanz-Matrix handelt, die mit der euklidischen Distanz zwischen den einzelnen Objekten gefüllt wird. Der `daisy`-Befehl nimmt dabei die folgenden Parameter entgegen. Zuerst wird die Datentabelle mit den Attributen `wg` angegeben. Dann wird die Art der Berechnung der Distanzmatrix mit dem Parameter `metric = "euclidean"` bestimmt. Damit ist die euklidische Distanz gemeint. Mit dem Parameter `stand=TRUE` wird eine Standardisierung der Werte vorgenommen. Dies ist nötig, weil sonst hohe Werte, wie die Angabe der Koordinaten, die eine Größenordnung von 10^6 (für die x-Koordinate) bis 10^7 (für die y-Koordinate) einen viel größeren Einfluss auf die Clusterbildung hätten, als z. B. Werte der Variable für die Grundflächenzahl, die sich nach ihrer Definition nur im Bereich $0 \dots 1$ bewegt. Durch die Standardisierung wird von den Werten aller Variable zuerst ihr Mittelwert abgezogen und dann durch die mittlere absolute Abweichung geteilt (R Core Team 2017). Die Folge ist, dass sich die Werte aller Variablen in der Größenordnung gleichen und der Einfluss auf die Bildung der Cluster dementsprechend gleich groß ist.

Nachdem die Abstandsmatrix erzeugt ist, kann nun die eigentliche Zuordnung zu den Clustern geschehen. Weil die Berechnung in der FME automatisch erfolgen soll, wurde hier auf das Paket `NbClust` zurückgegriffen, da es die Möglichkeit einer *Stopping Rule* zur Verfügung stellt, mit der die optimale Anzahl an Cluster durch ein statistisches Kriterium definiert wird. Die zu übergebenden

```

1 require(dplyr)
2 require(cluster)
3 require(NbClust)
4 wg <- select(fmeInput, grz_fme, gfz_fme, varatio_fme, +
5   avratio_fme, cpx_fme, cpy_fme)
6 dist.euclid <- daisy(wg, metric="euclidean", stand=TRUE)
7 res <- NbClust(data = wg, diss = dist.euclid, +
8   distance = NULL, min.nc = 2, max.nc = 15, +
9   method = "ward.D", index = "ch")
10 cluster <- res$Best.partition
11 fmeOutput<- cbind(fmeInput, cluster)

```

Abbildung 4.14: R-Script zur Clusteranalyse

Parameter sind die folgenden: mit dem Parameter `data = wg` wird der ursprüngliche Datensatz an den Algorithmus übergeben, der Parameter `diss = dist.euclid` gibt die zu verwendende Distanzmatrix an. Da diese schon vorliegt, muss der Algorithmus diese nicht selber berechnen, was mit dem Parameter `distance = NULL` auch unterbunden wird. Mit den Parametern `min.nc` und `max.nc` werden die minimale und die maximale Anzahl an gewünschten Clustern vorgegeben, hier 2 bis 15. Als Methode wird die Ward-Methode genutzt, aufgrund der einfachen euklidischen Distanz die entsprechende: `method = "ward.D"`. Die Indizierung und damit die *Stopping Rule* geschieht mit der Methode nach Calinski/Harabasz, siehe Abschnitt 2.4.3 *Clusteranalyse*, und wird mit der Option `index = "ch"` erwirkt. (R Core Team 2017)

Die Ergebnisse des Algorithmus werden an das Objekt `res` übergeben. Diese beinhaltet die Eigenschaft `res$Best.partition`, also die Angabe der Zuweisungen zur besten Clusterlösung. Diese werden im vorletzten Schritt in Zeile 10 an das Objekt `cluster` übergeben, das mit der Funktion `cbind` an den ursprünglichen Input in den `RCaller`-Transformer angehängt wird und im letzten Schritt in Zeile 11 zusammen mit diesem über den Output-Port des Transformers ausgegeben wird.

Wie schon erwähnt, werden die Ergebnisse der Clusteranalyse mit dem Transformer des Typs `FeatureMerger` wieder mit dem kompletten Datensatz vereint. Im letzten Schritt werden die Features durch die Transformer des Typs `FeatureColorSetter` eingefärbt, so dass bei der Anzeige im *FME Data Inspector* die Cluster sich voneinander farblich abheben. Damit lässt sich das Ergebnis direkt beurteilen. Für die Verwendung in R werden die Daten zusätzlich in einer csv-Datei, für den Import in QGIS in einer gml-Datei gespeichert.

5. Ergebnisse und Diskussion

5.1. Ergebnis der Literaturrecherche

In Unterkapitel 2.1 wurde eine Literaturrecherche zur Durchführung von Wasserbedarfsprognosen vorgenommen. Dazu lässt sich abschließend sagen, dass ein großer Aufwand betrieben wird, mit Hilfe von unterschiedlichen Einflussfaktoren Wasserbedarfsprognosen zu erstellen. Dabei ist der Verbraucher der wichtigste Faktor. Für die Berechnungen nach Fritsch et al. (2014) sind sehr detaillierte Daten nötig, obwohl hier zur Berechnung der Pro-Kopf-Bedarf nach DVGW (2008) herangezogen wird. Koegst und Tränckner (2010) machen aber deutlich, dass der Wasserbedarf nicht festgelegt werden kann, da er lokal stark variiert, was die Notwendigkeit eines konkreten Ansatzes zum Pro-Kopf-Verbrauch unterstreicht. Daher ist es notwendig, durch eine vorhersagbare Bevölkerungsentwicklung und einen auf lokalen sozioökonomischen Faktoren beruhenden Pro-Kopf-Verbrauch zurückzugreifen. Für die Erfassung von Einwohnerdaten werden u. a. persönliche Befragungen von Nutzern zu ihren Gewohnheiten durchgeführt. Dies wird von Grossmann und Hofmann (2008) sowie Kluge, Deffner et al. (2008) für das Prognose-Modell für Hamburg durchgeführt. Auch Schumann und Korth (2014) bzw. Martin, Korth und Schubert (2017) führen solche Befragungen durch, um Korrelationen zwischen dem maximalen Tages- und Spitzenverbrauch und der Zahl der Wohneinheiten und ihrer Bewohner zu bestätigen respektive zu widerlegen. Zusätzlich wurden im Rahmen ihres Forschungsprojektes die vorhandenen Hauswasserzähler durch Ultraschallzähler ersetzt. Dies bedeutet einen erheblichen Kosten-, Personal- und Zeitaufwand. Vor allem bei den Modellen, die auf einen Pro-Kopf-Bedarf angewiesen sind, ist die genaue Erfassung von Einwohnerzahlen immanent wichtig. Es werden zwar strukturelle Faktoren in die Modelle eingearbeitet, diese decken aber nicht gleichzeitig die potentiellen Einwohnerdichten ab, sondern werden z. B. für den Einfluss von Gartenbewässerung auf den Wasserbedarf herangezogen. Die Erfassung der Einwohner ist nicht allein ausschlaggebend, da laut Neunteufel, Richard, Perfler et al. (2012) die Unterschiede im Wasserverbrauch zwischen ländlichen und städtischen Versorgungsgebieten generell im jeweiligen Anteil des Gewerbes und der mitversorgten Industrie sowie öffentlichen Einrichtungen begründet liegen. Dies sehen Kluge, Deffner et al. (2008) ähnlich und halten daher eine Unterteilung nach städtischen und ländlichen Gebieten oder privaten und gewerblichen Verbrauchern für sinnvoll, was die Autoren auch umsetzen.

Die klimatischen Veränderungen haben in der Regel keine Auswirkungen auf den durchschnittlichen Wasserbedarf. Die grundlegende Dimensionierung der Wasserinfrastruktur muss allerdings mit erhöhten Spitzenwerten zurecht kommen. Demgegenüber stehen die Empfehlungen aus den diversen Forschungsvorhaben aktuell nicht (mehr) auf einen Wachstumstrend in der Bevölkerung zu setzen, sondern den Fokus mehr auf Wanderbewegungen, wie die Landflucht bzw. den anhaltenden Trend zur Suburbanisierung (Schmitz 2004, vgl.) zu setzen.

Allerdings sind diese Bewegungen und Trends nicht willkürlich, wie die Literatur zur Raumordnung behauptet, sondern durch eben diese Raumordnung gesteuert. Da die Raumordnung als übergeordnetes Werkzeug bislang von einem Bevölkerungswachstum ausging, wird dies in den Planungen auch weiterhin so fortgeführt werden. Das Problem für die Planung besteht darin, dass der Prognosezeitraum der Bevölkerungsentwicklung näher in der Zukunft liegt als der Zielhorizont für die Raumplanung. Bisher wird Raumplanung so betrieben, dass der Zielhorizont nicht zeitlich festgelegt wird, so dass eine eigentliche Evaluierung vom Gesamtkonzept her nicht durchgeführt werden kann (Schmitz 2004, S. 134). Es ist also nicht davon auszugehen, dass sich an dieser Praxis in naher Zukunft etwas ändert. Zusätzlich beeinflusst die Steuerungswirkung von Einzelnen (Investoren, Bauträgern) mindestens mittelbar die Planungen der Raumordnung.

5.2. Diskussion über die Datengrundlage

Die Datengrundlage ist im Fall des Bestandsdatenauszugs gut durchdacht und ordentlich gepflegt. Dies ist dadurch zu begründen, dass diese Datengrundlage die Umsetzung einer EU-Richtlinie darstellt. Die Komplexität dieser Datengrundlage, in der mehr als 100 Klassen vorliegen, deren Objekte auf unterschiedliche Art miteinander verknüpft sein können, macht eine vertiefende Einarbeitung in die Struktur nötig, um überhaupt mit dieser Datengrundlage arbeiten zu können.

Im Fall der 3D-Gebäudedaten ist zu sagen, dass die Datengrundlage stark von der amtlichen Datengrundlage des Bestandsdatenauszugs abweicht. Das mag darin begründet sein, dass die 3D-Gebäudedaten nicht auf dem Stand des Bestandsdatenauszugs sind. Dies ist nachvollziehbar, können doch nicht bei jeder Änderung (Abriss, Umbau, Neubau) neue Erfassungen der Daten durch Überflug und Laserscanning erfolgen. Die Struktur der Daten ist insgesamt gut gepflegt, was auch auf das verwendete Format zurückzuführen ist.

Die Daten der Bodenrichtwerte, die nur eine erste Grundlage für Verkäufe von Grundstücken darstellen, sind im Vergleich zu den vorangegangenen Datenquellen weniger gut gepflegt. Wie in 4.2 beschrieben, wurden hier falsche Datentypen der Variablen gesetzt, so dass diese für den Menschen zwar lesbar, aber von Rechnern nicht ohne weiteres verarbeitet werden können.

Die Zensus-Daten sind in dieser Arbeit auf die zu betrachtenden Gebiete reduziert worden, dies hat aber einen erheblichen Rechenaufwand bedeutet, da zu Beginn alle Rasterpunkte in der Bundesrepublik Deutschland eingelesen werden mussten. Hier empfiehlt es sich, für die einzelnen Bundesländer Datensätze bereitzustellen. Das hat den Vorteil, dass genauere Koordinatensysteme für kleinere Gebiete verwendet werden können.

5.3. Ergebnisse und Diskussion der Datenaufbereitung

Z0 Reduktion der Zensus-Daten auf das Zielgebiet Das Ergebnis für die Gemeinde Alpen der Datenreduktion der Zensus-Daten ist in Abb. 5.1 dargestellt. Die Ergebnisse für Lemgo und Solingen befinden sich in Anhang A. Auf der Karte sind die Gemeindegrenze und die Rasterzellen mit den Einwohnerdichten zu sehen, die am 200 m breiten Puffer um die Grenze abgeschnitten sind. Deutlich erkennbar ist die hohe Zahl an Rasterzellen ohne Angabe von Einwohnern (Wert: -1). Dies ist der Tatsache geschuldet, dass die Daten maskiert worden sind, vgl. Abschnitt 3.4. Die Häufungen an gültigen Werten stimmen mit den Koordinaten der Zentren der Stadtteile bzw. Siedlungsgebiete überein.

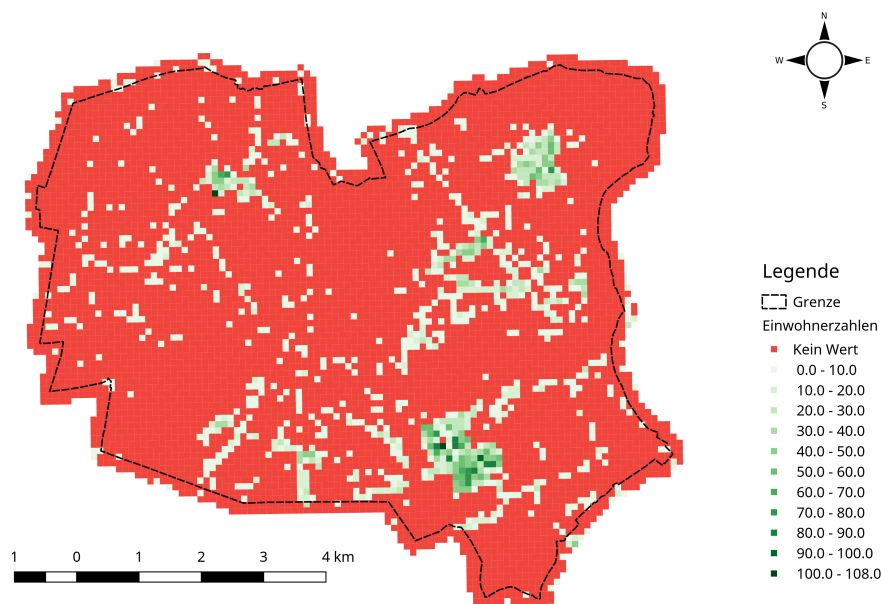


Abbildung 5.1: Ergebnis Z0 Reduktion der Zensus-Daten auf Zielgebiet, hier: Gemeinde Alpen

F0 Clippen der Flurstücke Das Ergebnis des Clippens der Flurstücke für die Gemeinde Alpen ist in Abb. 5.2 dargestellt. Auch hier ist die Grenze der Gemeinde eingezeichnet. Die Flurstücke schließen im Gegensatz zu den Zensus-Daten mit der Grenze exakt ab. Die Legende hat hier nur zwei Einträge, da keine weiteren Unterscheidungsmerkmale in diesem Schritt hinzugefügt worden sind.

F1 Aggregation der Flurstücke Das Ergebnis der Aggregation der Flurstücke für die Gemeinde Alpen ist in Abb. 5.3 dargestellt. In der Legende lässt sich ablesen, dass es Aggregationen von bis zu



Abbildung 5.2: Ergebnis *F0* Clippen der Flurstücke, hier: Gemeinde Alpen

sieben Flurstücken gibt, die unter derselben Buchungsnummer geführt werden. Die Mehrheit der Flurstücke wird aber nicht zusammengefasst.

F2 Projizieren der Zensus-Daten auf Flurstücke Das Ergebnis des Projizierens der Zensus-Daten auf die Flurstücke für die Gemeinde Alpen ist in Abb. 5.4 dargestellt. Dort, wo die Rasterzellen der Zensus-Daten die höchste Einwohnerdichte aufweisen, werden auch den Flurstücken die höchsten Werte zugewiesen. Auch hier lassen sich wie im Ergebnis von *Z0* die Zentren der Siedlungsgebiete gut erkennen.

F3 Verheiraten Flurstücke und Nutzungsart Das Ergebnis des Verheiratens der Flurstücke und der Nutzungsart für die Gemeinde Alpen ist in Abb. 5.5 zu sehen. Den Flurstücken werden hier die Nutzungsarten Wohngebiet, Mischgebiet, Verkehrsfläche und sonstige Fläche zugewiesen.

F4 Projizieren der Bodenrichtwerte auf Flurstücke Das Ergebnis der Projektion der BRW auf die Flurstücke in der Gemeinde Alpen ist in Abb. 5.6 dargestellt. Die Flurstücke sind nach Höhe des BRW weniger oder stärker eingefärbt. Auch hier sind die Siedlungszentren durch höhere Werte gekennzeichnet.

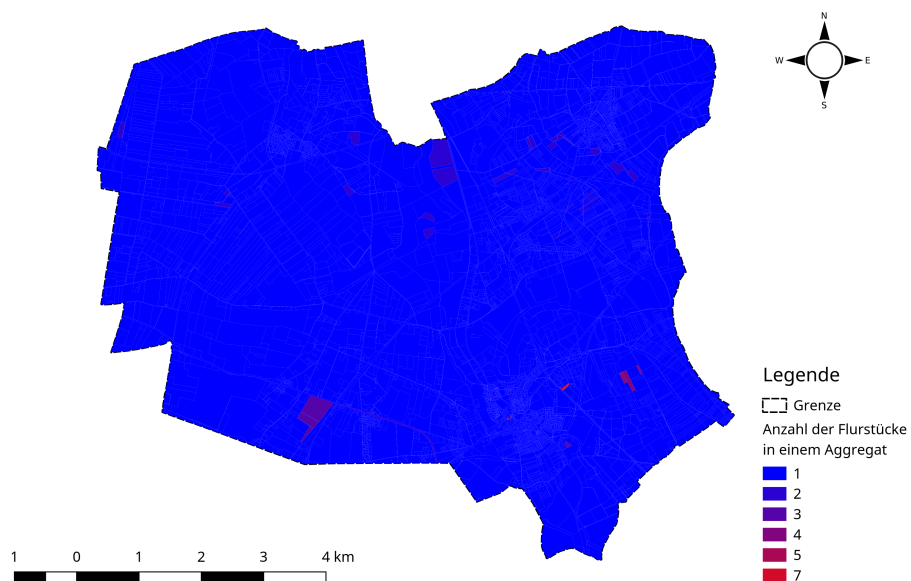


Abbildung 5.3: Ergebnis *F1* Aggregation der Flurstücke, hier: Gemeinde Alpen

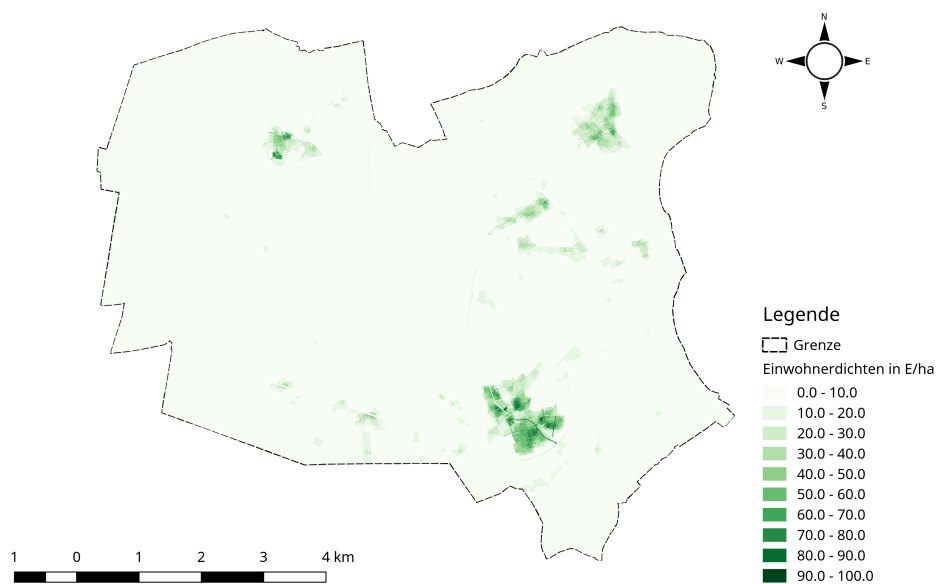


Abbildung 5.4: Ergebnis *F2* Projizieren der Zensus-Daten auf Flurstücke, hier: Gemeinde Alpen

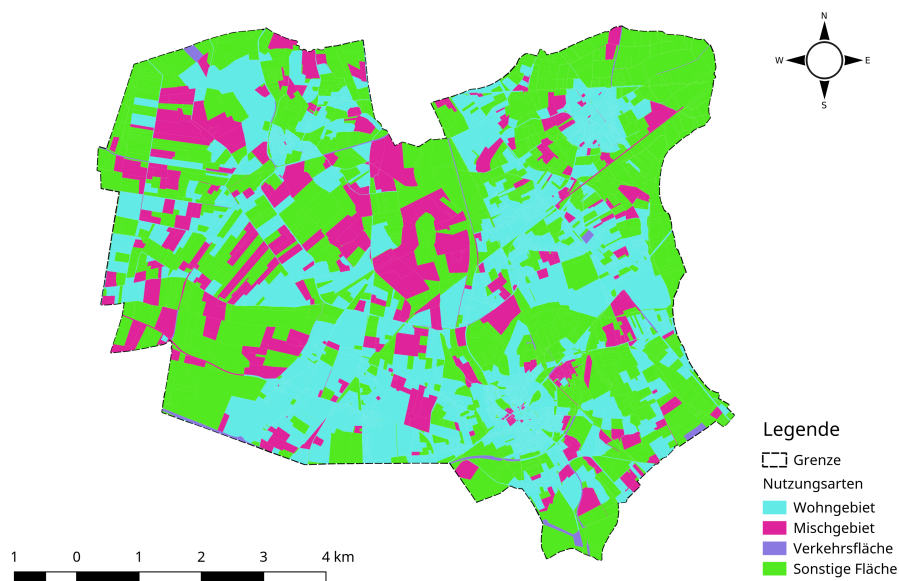


Abbildung 5.5: Ergebnis *F3* Verheiraten Flurstücke und Nutzungsart, hier: Gemeinde Alpen

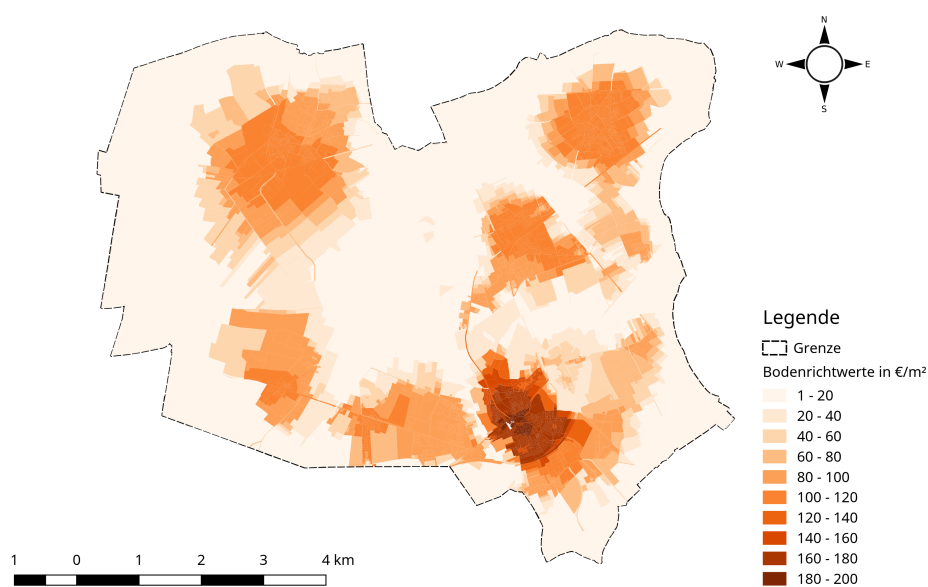


Abbildung 5.6: Ergebnis *F4* Projizieren der Bodenrichtwerte auf Flurstücke, hier: Gemeinde Alpen

G0/G1 Clippen der 3D-Gebäude und Reduktion der XML-Ebenen und Reduktion der 3D-Daten auf

2D Die an der Grenze ausgeschnittenen Gebäude in der Gemeinde Alpen sind in Abb. 5.7 zu sehen. Die aggregierten und nicht aggregierten Gebäude sind farblich gekennzeichnet. Die Gebäude sind bereits auf die zweidimensionale Ebene projiziert, die Geometriedaten wie Höhe, Außenfläche, Volumen und Grundfläche sind in den Daten enthalten. Auf eine Visualisierung dieser Werte wird an dieser Stelle verzichtet, da sie lediglich einen Zwischenschritt zur Berechnung der Siedlungskennzahlen darstellen.

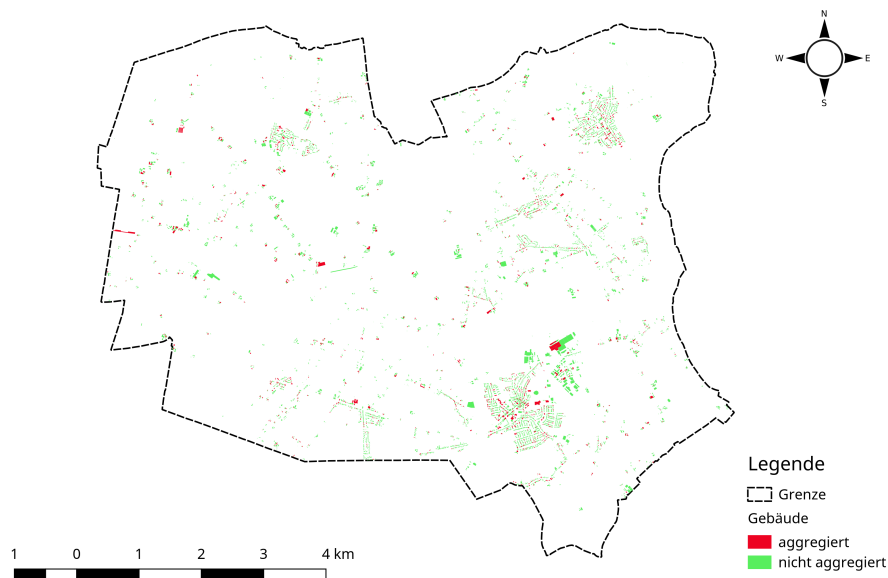


Abbildung 5.7: Ergebnis G0 Clippen der 3D-Gebäude und Reduktion der XML-Ebenen, hier: Gemeinde Alpen

G2 Verheiraten von Gebäudedaten Um zu prüfen, ob ein in den 3D-Daten vorhandenes Gebäude amtlich geführt ist, wird dieses mit den Daten aus dem Bestandsdatenauszug verglichen. Gebäude, die nicht amtlich vermerkt sind, werden mit diesem Schritt eliminiert. Das Ergebnis ist in Abb. 5.8 dargestellt.

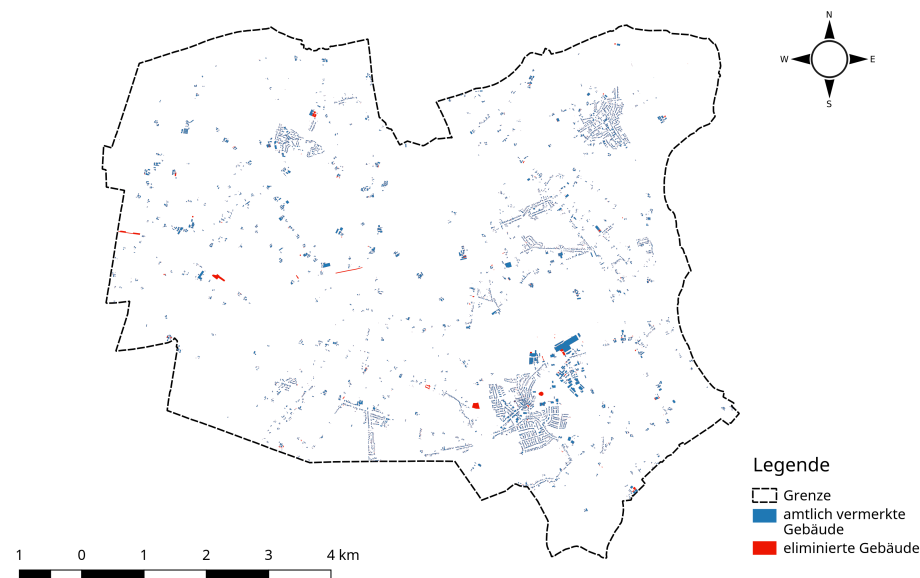


Abbildung 5.8: Ergebnis G2 Verheiraten von Gebäudedaten, hier: Gemeinde Alpen

GF Verheiraten der Daten Im letzten Schritt der Datensynthese werden die Daten der Gebäude und der Flurstücke miteinander verheiratet. Dies ermöglicht, die Siedlungskennzahlen zu berechnen. Die robusteste und zugleich genaueste Kennzahl ist die GRZ. Ihre Verteilung von 0 bis 1 ist in Abb. 5.9 dargestellt. Deutlich zu sehen ist, dass die Werte in den den Zentren der Siedlungsgebiete deutlich höher sind als an deren Rändern. Bei der GFZ ist zu erkennen, dass die Höhe der Gebäude Einfluss auf die Bildung dieser Kennzahl hat. Gebiete mit höherer Bebauung sind deutlich abgegrenzt, siehe Abb. 5.10. Auch die BMZ ist in den Zentren der Siedlungsgebiete deutlich höher, wie in Abb. 5.11 zu sehen. Bei der Kompaktheit ist keine Abhebung in den Zentren der Siedlungsgebiete zu beobachten, siehe Abb. 5.12. Dies liegt an der ungenauen Berechnung der Außenflächen durch FME. Es lässt sich schon an dieser Stelle ein Zusammenhang zwischen den Siedlungskennzahlen (mit Ausnahme der Kompaktheit), den Bodenrichtwerten und den Einwohnerdichten vermuten. Dieser Zusammenhang wird mit der Clusteranalyse statistisch überprüft. Die Ergebnisse sind in Abschnitt 5.4 aufgelistet.

Berechnung der Außenflächen der Gebäude Die Schwierigkeit bei der Berechnung der diversen Werte für die Gebäude, wie Volumen, Außenflächen und Höhe der Gebäude, besteht darin, dass die XML-Struktur der Gebäudedaten eine Hierarchie aufweist. Komplexere Gebäude werden bei der Erstellung der 3D-Daten durch das Land Nordrhein-Westfalen in Gebäudeteile aufgeteilt. Dies ist dann sinnvoll, wenn die Daten beispielsweise über das Internet dynamisch geladen werden und zuerst

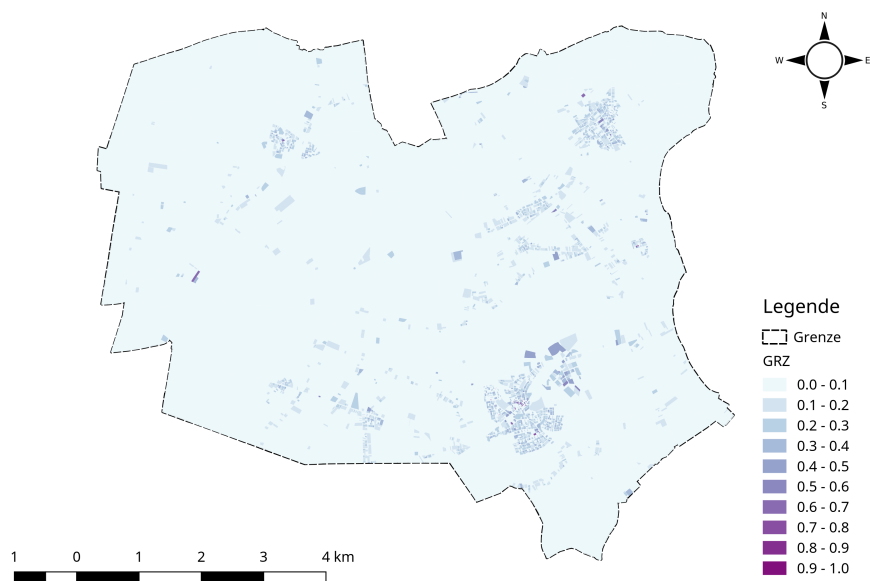


Abbildung 5.9: Ergebnis *GF* Verheiraten der Daten - Grundflächenzahl, hier: Gemeinde Alpen

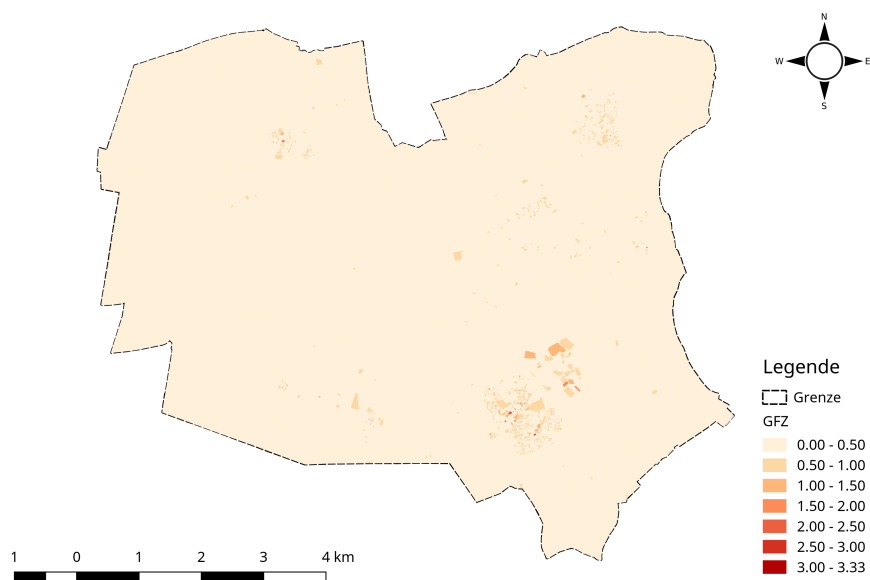


Abbildung 5.10: Ergebnis *GF* Verheiraten der Daten - Geschossflächenzahl, hier: Gemeinde Alpen

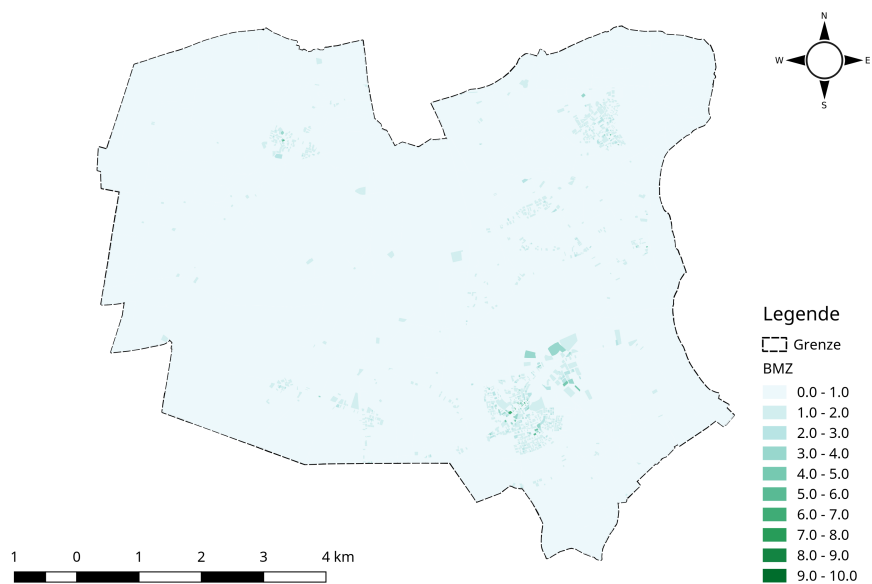


Abbildung 5.11: Ergebnis *GF* Verheiraten der Daten - Baumassenzahl, hier: Gemeinde Alpen

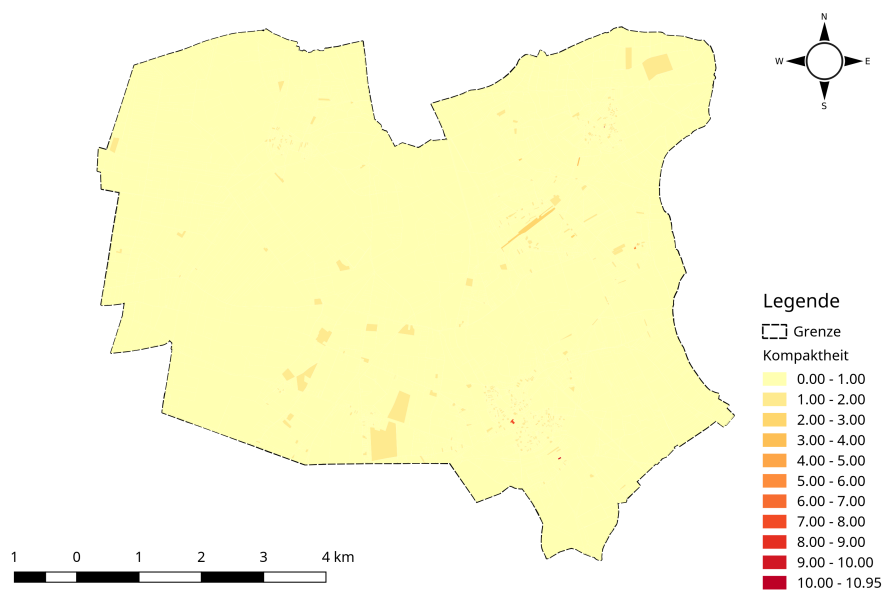


Abbildung 5.12: Ergebnis *GF* Verheiraten der Daten - Kompaktheit, hier: Gemeinde Alpen

die Informationen über die Existenz eines Gebäudes übertragen und auf Anforderung danach die Details wie die Geometriedaten übertragen werden. Weil XML und seine Abkömmlinge wie CityGML, GML und INSPIRE XML für Anwendungen im Internet konzipiert worden sind, ist eben diese innere Hierarchie für sie wesentlich. Im vorliegenden Anwendungsfall macht die Unterteilung in Gebäudeteile Probleme bei der Berechnung, speziell bei der Berechnung der Außenflächen eines Gebäudes. FME wertet solche Flächen, die zwischen zwei Gebäudeteilen der 3D-Datengrundlage liegen, auch als Außenflächen. Es ist deswegen erforderlich, die Gebäudeteile in ihre Außenflächen zu zerlegen, um sie daraufhin zu einem gesamten Gebäude zusammen zu rechnen. Damit wird sichergestellt, dass innere Flächen eliminiert werden und nicht mehr in die Berechnung der Außenflächen einfließen können. Die Unterschiede in der Flächenberechnung zwischen den Rohdaten des Landes Nordrhein-Westfalen und den hier aggregierten Daten soll im Folgenden anhand eines Beispiels verdeutlicht werden.

Als Beispiel wird ein Gebäude in der Gemeinde Alpen herangezogen, das in der Datengrundlage aus zwei Teilen besteht. Es hat die GML-Identifikation DENW34AL0000aTy1. Die beiden Gebäudeteile grenzen an einer Seite aneinander, siehe Abb. 5.13. In diesem Screenshot aus dem FME *Data Inspector* ist ein Gebäudeteil (im Hintergrund) markiert und daher sind die Umrandungslinien der einzelnen Polygone rot dargestellt. Das vordere Objekt ist nicht markiert und daher durchgängig grau dargestellt.

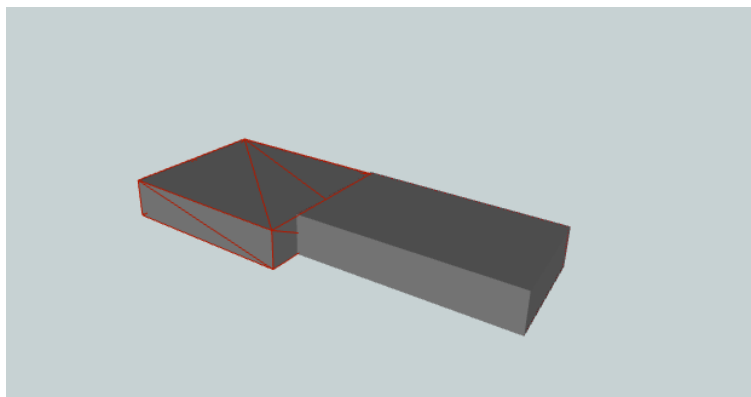


Abbildung 5.13: Beispiel Berechnung Außenfläche: Gebäude aus zwei Teilen

Zur Überprüfung der Berechnung durch FME wird zunächst eine manuelle Berechnung der gesamten Fläche A_G inklusive der Innenflächen durchgeführt. Dies erfolgt über die Summe der einzelnen Teilflächen, wobei A_n eine Teilfläche darstellt und N die Gesamtzahl der Flächen:

$$A_G = \sum_{n=1}^N A_n \quad (5.1)$$

Die Berechnung erfolgt mit einer Tabellenkalkulation und ist in Tabelle 5.1 dargestellt. Die Anzahl der

Flächen beträgt in diesem Fall 14. Dies ist auffällig, weil ein Quader per Definition nur 6 Außenflächen besitzt. Hier sind aber zwei Außenflächen des breiteren Gebäudes in zwei Teilflächen aufgeteilt worden, was die Gesamtzahl von $2 \cdot 6 + 2 = 14$ Flächen erklärt. Dabei ist die Fläche mit der Nummer 13 die Fläche des schmaleren Gebäudeteils und grenzt an das breitere Gebäudeteil an. Diese darf somit nicht in die Berechnung einfließen und wird gestrichen. Die Grenzfläche des größeren Gebäudes ist in den Rohdaten unterteilt in Fläche 3 und 4. Die Summe der beiden Flächen muss also um die Fläche mit der Nummer 13 reduziert werden, um keine Fläche doppelt in die Berechnung einfließen zu lassen. Die Summe der Flächen mit den Nummern 3 und 4 beträgt $80,71 \text{ m}^2$. Abzüglich der Fläche 13 ergibt sich ein Wert von $14,05 \text{ m}^2$. Dieser Wert wird für die Fläche Nummer 3 eingesetzt. Der Wert von Fläche 4 wird auf null gesetzt. Die korrekte Größe der Außenfläche beträgt somit $2011,16 \text{ m}^2$.

Flächen-Nr.	Größe in m^2	reduzierte Größe in m^2
1	365,7133074993	365,7133074993
2	365,7133074993	365,7133074993
3	40,3554550872	14,0477504951
4	40,3554550873	0
5	90,87550026	90,87550026
6	39,1191788949	39,1191788949
7	39,119784516	39,119784516
8	90,2240392089	90,2240392089
9	353,652195	353,652195
10	353,652195	353,652195
11	66,1727450966	66,1727450966
12	118,5882985068	118,5882985068
13	66,6631596794	0
14	114,236940804	114,236940804
Σ	2144,4415621397	2011,1152427809

Tabelle 5.1: Manuelle Berechnung der Außenfläche eines Gebäudes

Die von FME berechnete Außenfläche des gesamten Gebäudes beträgt inklusive der Grundfläche $2144,4 \text{ m}^2$. Für die Berechnung der Kompaktheit wird die Grundfläche abgezogen, was für dieses Beispiel keine Rolle spielt, da hier die Gesamtfläche verglichen wird.

An dieser Stelle stößt die Software FME an ihre Grenzen. Die Flächenberechnung bei dreidimensionalen Objekten ist nicht ausgereift. Zwar gibt es Ansätze in der Softwareentwicklung, die die Identifizierung von Wänden, Decken und Böden eines einfachen Objektes als Grundlage für die Zusammenlegung von Innenwänden erlauben. Dafür stellt die Firma *SafeSoftware* online einen Transformer zur Verfügung. Dieser steckt allerdings noch in der Testphase. Der Transformer trägt den Namen `BRepSolidBoundaryCreator.Beta`, wobei die Endung deutlich macht, dass dieser sich noch in der Beta-Phase befindet. Er ist im aktuellen Entwicklungsstand noch nicht in der Lage, zufriedenstellende Ergebnisse zu produzieren. Zur angedachten Funktionsweise sei gesagt, dass der Transformer über seinen Input-Port Features aufnehmen soll, die Oberflächen eines dreidimensionalen Objektes

darstellen, um über seine drei Output-Ports diese auszugeben, und zwar als Top (Oberseite), Side (Seite) und Bottom (Unterseite). In der Praxis hat sich allerdings gezeigt, dass dieser Transformer nicht in der Lage ist, die dreidimensionalen Gebäudedaten zu verarbeiten. Manche Flächen werden falsch eingeordnet, andere werden direkt über den Rejected-Port abgewiesen. Die wenigsten Oberflächen-Features werden korrekt verarbeitet.

Generell stellt sich die Frage, ob es Sinn ergibt, die auf einem Grundstück vorhandenen und einander angrenzenden Gebäude bzw. deren Außenflächen um eben diese innenliegenden Wände zu reduzieren, da es auch den Fall geben kann, dass sich zwei Gebäude auf zwei getrennten Grundstücken, sich innenliegende Wände teilen können, z. B. bei einer Häuserzeile. Es gibt zwei Möglichkeiten: gemeinsame Innenwände werden entweder ignoriert oder es muss eine individuelle Fallentscheidung getroffen werden. Letzteres macht aufgrund der Automatisierung von FME keinen Sinn, da manuell eingegriffen werden müsste. Es wird daher entschieden, dass aufgrund der fehlerhaften Berechnung der Außenfläche die Kompaktheit nicht zur Clusteranalyse herangezogen werden kann.

5.4. Ergebnisse und Diskussion der Datenanalyse

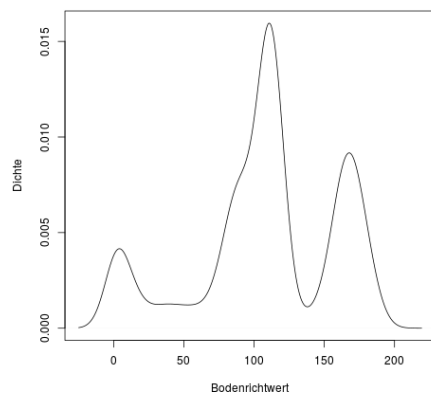
Zuerst werden wie in 2.4.1 beschrieben univariate Häufigkeitsverteilungen vorgenommen, um einen ersten Überblick über die Verteilungen der Ausprägungen zu erhalten. Für die Visualisierung wird ein Kern-Dichte-Schätzer genutzt. Die Verteilungen für den BRW, die GRZ, die GFZ, die BMZ und die Einwohnerdichte sind für die Gemeinde Alpen in Abb. 5.14 dargestellt. Die Berechnung erfolgt nach dem Satz von Nadaraya mit einem Gaußkern. Die Anzahl der Datensätze beträgt in jedem Fall $N = 3368$. Dies entspricht der Anzahl der Flurstücke in den Wohngebieten. Die Bandbreiten oder -weiten sind die folgenden: Bodenrichtwert 8,794, Einwohnerdichte 3,447, GRZ 0,02088, GFZ 0,04304 und für die BMZ 0,1282. Die Bodenrichtwerte zeigen drei Spitzen, siehe Abb. 5.14a. Die Verteilung ist in diesem Fall trimodal. Die Verteilungen der Ausprägungen der übrigen Variablen können dagegen als nahezu unimodal bezeichnet werden.

Durch die Clusteranalyse nach dem Ward-Verfahren werden die Cluster generiert. In die Clusteranalyse werden alle berechneten Siedlungskennzahlen mit Ausnahme der Kompaktheit sowie die Koordinaten der Flurstücke miteinbezogen. Da eine Clusteranalyse durch topologische Zusammenhänge im statistischen Sinne (Distanzmaße) durchgeführt wird und im vorliegenden Fall auch die Topologie im geometrischen Sinne durch den Einfluss der Koordinaten einbezogen wird, kann man in diesem Fall den Begriff einer topologisch-topologischen Analyse einführen.

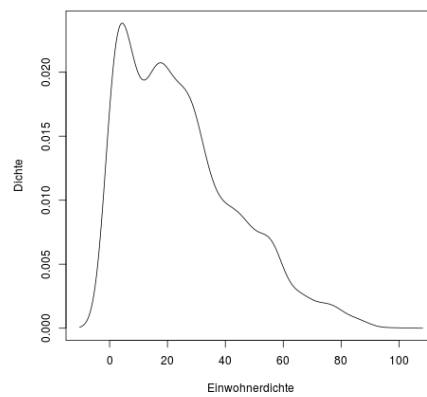
Dabei wird ebenfalls die optimale Clusterzahl nach Calinski/Harabasz (vgl. Backhaus et al. 2016, S. 496) bestimmt. Diese beträgt im Fall der Gemeinde Alpen 8. Im Vergleich zu den beiden anderen Städten ist die Anzahl höher, was daran liegt, dass in der Gemeinde Alpen mehrere Siedlungszentren vorliegen. Durch die Einbeziehung der Koordinaten kann der Algorithmus dies berücksichtigen. In Abb. 5.15 sind die gebildeten Cluster für die Wohngebiete der Gemeinde Alpen dargestellt.

Die Verteilungen der Ausprägungen der Variablen werden mit einem Boxplot grafisch visualisiert. Für je eine Variable werden die Verteilungen in den einzelnen Clustern gegenübergestellt. Diese sind in Abb. 5.16 für die verwendeten Variablen dargestellt. Die Bodenrichtwerte unterscheiden sich deutlich in den Clustern, die statistischen Kennzahlen sind in Tabelle 5.2 gelistet. Dabei entspricht der Median (50% Quartil) dem Strich innerhalb der Box, das untere Quartil (25 % Quartil) der unteren Grenze der Box und das obere Quartil (75 % Quartil) der oberen Grenze. Das Minimum wird als 0%-Quartil bezeichnet. Das 100%-Quartil ist das Maximum.

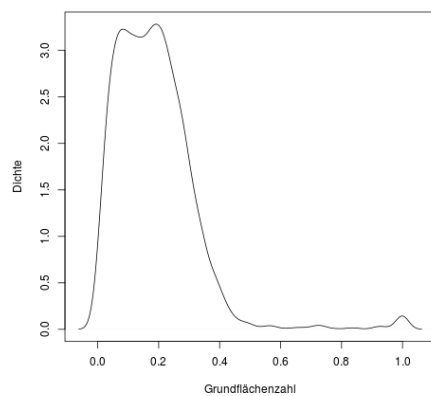
Im Nordwesten der Gemeinde Alpen, um den Stadtteil Veen, haben sich die Cluster 1 und 8 herausgebildet. Diese unterscheiden sich voneinander deutlich in ihrem Bodenrichtwert. Während dieser in Cluster 1 stark variiert, das 25%-Quartil liegt bei $2,17 \text{ €/m}^2$, das 75%-Quartil liegt bei $89,07 \text{ €/m}^2$, der Median bei $38,61 \text{ €/m}^2$, ist er in Cluster 8 stark eingegrenzt, dafür aber deutlich höher, siehe Abb. 5.16a. Hier liegen die Werte für das 25%-Quartil bei $108,28 \text{ €/m}^2$, für den Median bei $109,48 \text{ €/m}^2$ und das obere Quartil bei $109,92 \text{ €/m}^2$. Auch die Einwohnerdichten sind stark unterschiedlich: während im Cluster



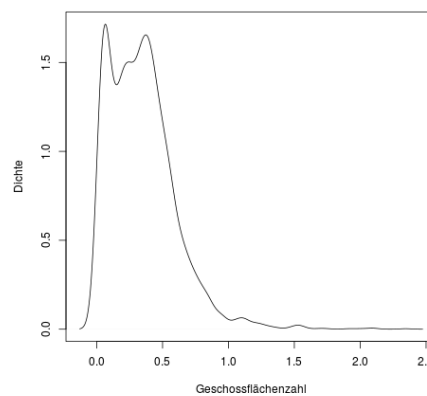
(a) Bodenrichtwert



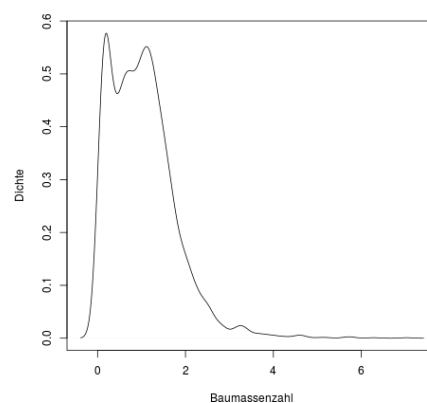
(b) Einwohnerdichte



(c) Grundflächenzahl



(d) Geschossflächenzahl



(e) Baumassenzahl

Abbildung 5.14: Kern-Dichte-Schätzer, hier: Gemeinde Alpen

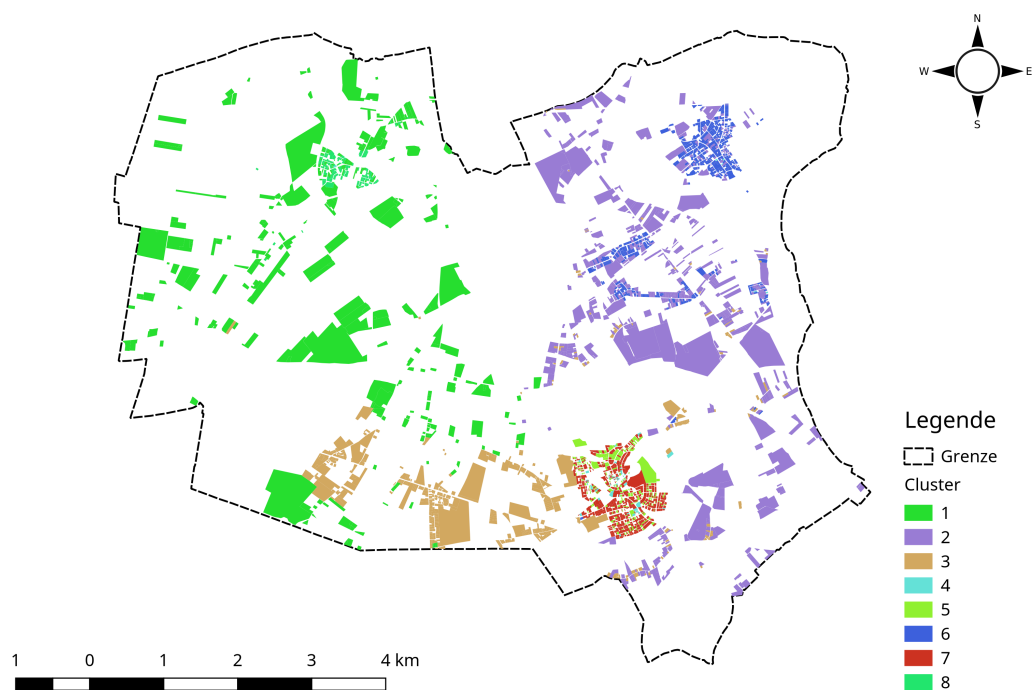


Abbildung 5.15: Ergebnis GF Clusteranalyse, hier: Gemeinde Alpen

1 der Median bei 3,75 E/ha liegt, ist er in Cluster 8 mit 27,55 E/ha deutlich höher. Dort leben mehr Menschen, was durch die in Abb. 5.15 abgebildete vorhandene Siedlungsstruktur zu vermuten war. Auch in der Bebauung sind die Unterschiede erkennbar, was durch die Ausprägung von GRZ, GFZ und BMZ zu erkennen ist.

Im Nordosten liegt der Stadtteil Menzelen. Hier dominieren die Cluster 2 und 6. Auch sie unterscheiden sich sehr stark im Bodenrichtwert: während in Cluster 2 die Spannweite des BRW sehr groß ist (unteres Quartil bei 20,15 €/m², oberes Quartil bei 112,14 €/m²), ist bei Cluster 6 die Streuung um den Median, der bei 113,76 €/m² liegt, sehr gering. Allerdings gibt es hier einige Ausreißer nach unten, das 0%-Quartil liegt hier bei 41,96 €/m². Bezogen auf die Einwohnerdichte ist Cluster 2 deutlich dünner besiedelt, die Spannweiten überschneiden sich. Die Boxen sind dennoch deutlich voneinander abgegrenzt. So liegt die obere Grenze (oberes Quartil) bei Cluster 2 bei 19,15 E/ha, die untere Grenze (unteres Quartil) von Cluster 6 bei 20,35 E/ha. Ähnliches gilt für den Grad der Bebauung. Als Beispiel sei die Spannweite der GRZ genannt, wo zwar auch Überschneidungen zu sehen, die Boxen dennoch unterscheidbar sind. Bei Cluster 6 liegt das untere Quartil bei 0,12, bei Cluster 2 liegt das obere Quartil bei 0,18. Wie in Abb. 5.15 zu sehen ist, steht Cluster 6 für die Siedlungsgebiete und Cluster 2 für die ländlichen Gebiete.

In der Dichte der Bebauung dem Cluster 2 sehr ähnlich ist Cluster 3. Auch der Median des BRW befindet sich im selben Bereich. Lediglich die BMZ und die GFZ haben höhere Werte, was auf eine mehrgeschossige Bauweise hinweist. Die Einwohnerdichte bezogen auf die Grundfläche ist dabei dem Cluster 2 sehr ähnlich, allerdings steht den Einwohnern hier mehr Wohnfläche zur Verfügung. Für eine Übersicht aller Werte, vergleiche Tabelle 5.2.

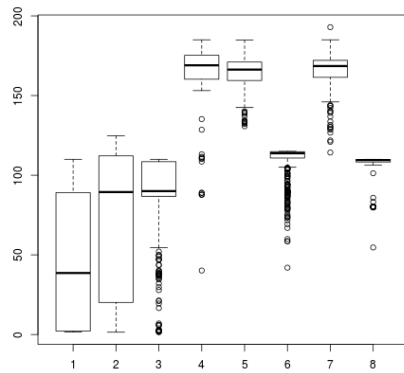
Cluster 4, 5 und 7 liegen im eigentlichen Ortskern von Alpen. Die Bodenrichtwerte sind hier nahezu identisch. Auch die Einwohnerdichten liegen im selben Bereich, wobei sich Cluster 4 ein wenig nach oben abhebt. Das Unterscheidungskriterium ist hier eindeutig die Dichte der Bebauung. Cluster 4 hat die höchste Bebauungsdichte, Cluster 5 die niedrigste und Cluster 7 liegt dazwischen. Am deutlichsten ist dies bei den Medianen der GRZ zu sehen: Cluster 5 mit 0,08, Cluster 7 mit 0,25 und Cluster 4 mit 0,61.

Die Ergebnisse für Lemgo und Solingen sind in Anhang B abgebildet. Da Lemgo und Alpen Gemeinden mit ländlichem Charakter darstellen, wurde die Clusteranalyse für Solingen auf das Zentrum beschränkt, so dass hier die Clusterbildung in einem Kerngebiet betrachtet werden kann. Die Clusteranalyse der anderen beiden Städte ergibt eine deutlich geringere Anzahl an Clustern als bei der Gemeinde Alpen. Dies liegt darin begründet, dass diese Städte im Vergleich zu Alpen nur ein Zentrum aufweisen.

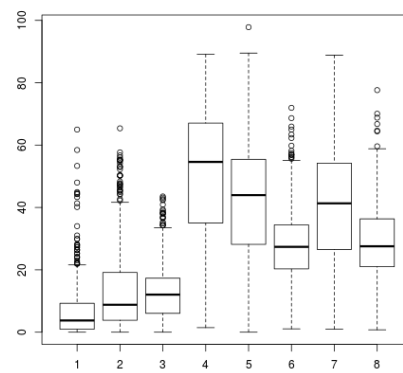
Im Fall von Lemgo werden 5 Cluster gebildet. Hierbei haben Cluster 2 bis 4 bezogen auf die Einwohnerdichte, den Bodenrichtwert und die Art der Bebauung keinen größeren Unterschied zueinander.

Cluster	Variable	0%-Quartil	25%-Quartil	50%-Quartil	75%-Quartil	100%-Quartil
1	Bodenrichtwert	1,500000	2,166667	38,617802	89,073154	109,999684
	Einwohnerdichte	0,000000	0,950606	3,752867	9,270316	64,977790
	Grundflächenzahl	0,0009778265	0,0371282715	0,0810282078	0,1211261007	0,3317772377
	Geschossflächenzahl	0,0009838694	0,0546635411	0,1243267767	0,2190259134	0,4684323740
	Baummassenzahl	0,002951608	0,161291390	0,372207662	0,652828424	1,355310842
2	Bodenrichtwert	1,46634	20,15484	89,41038	112,14327	124,76797
	Einwohnerdichte	0,000000	3,833555	8,789912	19,145774	65,376625
	Grundflächenzahl	0,002199782	0,051316978	0,086295269	0,122382832	0,344118569
	Geschossflächenzahl	0,0006730637	0,0662934890	0,1443098176	0,2169393156	0,5969660759
	Baummassenzahl	0,002019191	0,193456392	0,429541556	0,642874213	1,687217516
3	Bodenrichtwert	1,466667	86,765801	90,016623	108,474728	109,994304
	Einwohnerdichte	0,000000	6,045874	12,040701	17,315759	43,526726
	Grundflächenzahl	0,0001319527	0,1089184806	0,1669315684	0,2237472142	0,6383705643
	Geschossflächenzahl	0,0000,95968	0,1978488	0,3115230	0,4197608	0,9630408
	Baummassenzahl	0,0002879044	0,5843353680	0,9288464493	1,2536692303	2,8891224627
4	Bodenrichtwert	40,16746	160,28374	168,99373	175,27295	184,98821
	Einwohnerdichte	1,455183	35,015749	54,602231	67,021348	89,125018
	Grundflächenzahl	0,2477756	0,3885489	0,6075605	0,9871294	1,0000000
	Geschossflächenzahl	0,7285959	0,9350000	1,1017178	1,2609956	2,3445402
	Baummassenzahl	2,185788	2,776172	3,239433	3,759928	7,033621
5	Bodenrichtwert	130,7235	159,4567	166,2908	171,0015	184,8705
	Einwohnerdichte	0,00000	28,17210	43,95655	55,37698	97,81968
	Grundflächenzahl	0,005592631	0,055526717	0,076415973	0,108737399	0,205086665
	Geschossflächenzahl	0,002654635	0,048950520	0,072223223	0,157951324	0,293582656
	Baummassenzahl	0,007963906	0,146851559	0,216669670	0,473853972	0,864139892
6	Bodenrichtwert	41,9576	110,7722	113,7614	114,7670	114,9964
	Einwohnerdichte	1,06367	20,35169	27,35996	34,42571	71,95084
	Grundflächenzahl	0,1094923	0,1844704	0,2285094	0,2811443	0,5939159
	Geschossflächenzahl	0,1407151	0,3384019	0,4198409	0,5252944	1,1516378
	Baummassenzahl	0,4221453	1,0060083	1,2510836	1,5564662	3,4549133
7	Bodenrichtwert	114,2278	161,4331	168,5174	172,0666	192,9217
	Einwohnerdichte	0,9670207	26,5555549	41,3188165	54,1964905	88,8543753
	Grundflächenzahl	0,09694796	0,20973423	0,25413482	0,30305721	0,57113450
	Geschossflächenzahl	0,1894781	0,3975950	0,4929123	0,6234190	0,9680677
	Baummassenzahl	0,5684343	1,1861734	1,4620754	1,8389888	2,9042031
8	Bodenrichtwert	54,63425	108,28220	109,48406	109,92125	109,99337
	Einwohnerdichte	0,7610286	20,9866420	27,5453517	36,3459639	77,6257661
	Grundflächenzahl	0,09760325	0,19353880	0,23499648	0,28115855	0,48904859
	Geschossflächenzahl	0,2788958	0,3935120	0,4592596	0,5343186	1,2100771
	Baummassenzahl	0,8366874	1,1762023	1,3732532	1,6006531	3,6302314

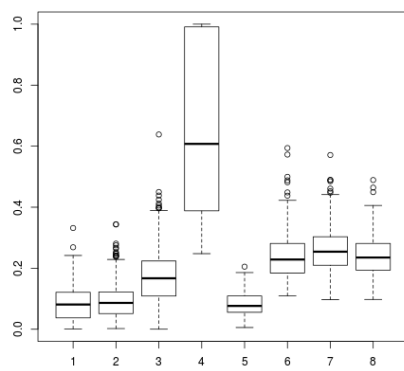
Tabelle 5.2: Statistische Kennzahlen der Cluster, hier: Gemeinde Alpen



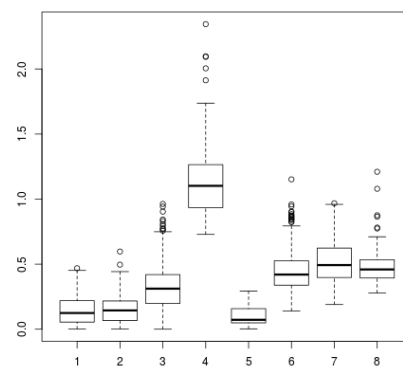
(a) Bodenrichtwert



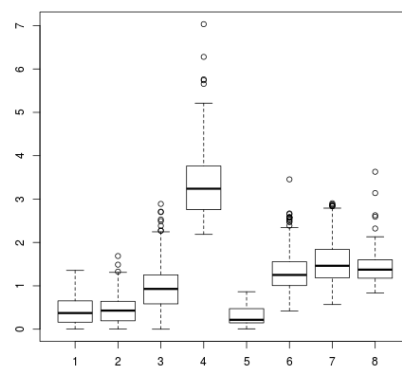
(b) Einwohnerdichte



(c) Grundflächenzahl



(d) Geschossflächenzahl



(e) Baummassenzahl

Abbildung 5.16: Boxplots zur Visualisierung der Ergebnisse, hier: Gemeinde Alpen

Die Objekte in diesen Clustern liegen am Stadtrand und weisen eine dünne Besiedlung auf, siehe Abb. B.1 im Anhang. Im Ortskern, der von den Clustern 1 und 5 eingenommen wird, gibt es starke Unterschiede. Die Bodenrichtwerte von Cluster 1, die auch das deutlichste Unterscheidungsmerkmal zu den vorher genannten Clustern darstellen, sind deutlich geringer als in Cluster 5. Dies gilt auch für die Einwohnerdichte. Auch bei der Bebauungsdichte liegt Cluster 1 deutlich unter Cluster 2. Die genauen Werte der statistischen Kennzahlen sind im Anhang in Tabelle B.1 hinterlegt. Die Boxplots für Lemgo sind in Abb. B.2 zu sehen.

Die Clusteranalyse für Solingen Mitte führt zu 5 Clustern. Diese sind in Abb. B.3 im Anhang dargestellt. Der Bodenrichtwert variiert zwischen den Clustern nur leicht. Er steigt mit zunehmender Clusterzahl an: Cluster 1 hat den geringsten Bodenrichtwert, Cluster 5 den höchsten. Auch die Einwohnerdichte bezogen auf die Grundstücksfläche ist fast konstant, siehe Abb. B.4. Dahingegen ist die Bebauungsdichte sehr unterschiedlich. Während in Cluster 1 der Median für die GRZ bei 0,16 liegt, ist er in Cluster 5 am höchsten mit knapp 0,6. Ähnliches gilt auch für die anderen Siedlungskennzahlen, siehe Tabelle B.2.

Im Vergleich untereinander hat sich gezeigt, dass die Bodenrichtwerte in Solingen Mitte am höchsten sind. Dabei sind sie kaum variabel, so dass die Clusterbildung nicht durch diese, auch nicht durch die Einwohnerdichte, die ebenfalls kaum variabel ist, sondern vielmehr durch die Dichte der Bebauung geschieht. In Lemgo und Alpen hat der Bodenrichtwert einen sehr viel größeren Einfluss auf die Clusterbildung. Auch die Einwohnerdichte hat an den beiden Orten einen höheren Einfluss auf die Clusterbildung. Das liegt daran, dass hier die Clusteranalyse bis an den Rand der Orte durchgeführt wurde, wo die Einwohnerdichte deutlich abfällt. Dies ist bei dem separat betrachteten Zentrum von Solingen nicht der Fall.

5.5. Fazit und Kritik

Die Datensynthese mit FME hat gezeigt, dass mit dem Interpolationsverfahren der invers quadratischen Gewichtung Bodenrichtwerte und Einwohnerdichten auf die Flurstücke projiziert werden können. Im Fall der Einwohnerdichten ist das Verfahren sehr robust, da die Dichte der Datenpunkte konstant ist. Im Fall der Bodenrichtwerte funktioniert das Verfahren ebenso, obwohl die Möglichkeit einer Häufung von Datenpunkten bestehen kann. Hier ist ein rein statistisches Verfahren, wie z. B. das *kriging*, bei vielen Datenpunkten die bessere Wahl, da es mit einer Häufung der Datenpunkte umgehen kann. Da aber ohnehin nur die drei benachbarten Datenpunkte für die invers quadratische Gewichtsfunktion zur Berechnung herangezogen werden, wird diese Einschränkung umgangen. Das Ergebnis zeigt, dass die Bodenrichtwerte gleichmäßig und ohne deutliche Sprünge auf die Flurstücke projiziert worden sind.

Die Aggregation der Flurstücke hat wesentliche Vorteile für die Verknüpfung von Gebäude- und Flurstückdaten. Wäre dieser Schritt nicht erfolgt, hätte die Zuweisung der Gebäude zu einem Flurstück Probleme bereitet, da es Überlappungen gegeben hätte. So ist sichergestellt, dass ein Gebäude eindeutig einem Flurstück zugeordnet werden kann.

Die vorbereitenden Datenreduktionen haben die Berechnungszeit wesentlich verkürzt. FME ist generell sehr ressourcenlastig. Bevor die Berechnungen und die Schreibvorgänge in die Ausgabedateien erfolgen und der gesamte Prozess abgeschlossen ist, werden alle Features im Arbeitsspeicher belassen, auch dann wenn sie gar nicht den Weg in die Ausgabedateien nehmen und verworfen werden.

Die Berechnung der Geometriedaten ist mit Ausnahme der Außenfläche ebenfalls erfolgreich. Hier ist eine Verbesserung seitens der Entwickler der Software wünschenswert. Da aber immerhin drei weitere Siedlungskennzahlen errechnet werden konnten, ist dies zu verkraften. Aus den 3D-Daten ist das Volumen in die BMZ eingeflossen und in die GFZ die Höhe der Gebäude. Dies ist ohne die Verwendung von 3D-Daten nicht möglich. Liegt die Datengrundlage nur zweidimensional vor, kann nur die GRZ berechnet werden, die bei der nachgeschalteten Clusteranalyse kein ausreichendes Unterscheidungsmerkmal ist.

Es hat sich gezeigt, dass durch eine Clusteranalyse Siedlungsgebiete mit ähnlichen siedlungsstrukturellen Kennzahlen, sozioökonomischen Faktoren, wie dem BRW, und der Einwohnerdichte zu Gruppen zusammengeschlossen werden können. Durch die Einbeziehung der Koordinaten können dabei die Zentren von Siedlungsgebieten sowie die Ähnlichkeiten von Bebauungsarten bezogen auf die Nachbarschaft zur Clusterbildung herangezogen werden.

Für die Auslegung von Trinkwasserinfrastrukturen hat die Einbeziehung der Koordinaten ebenfalls einen Vorteil. So ergibt es bei der Auslegung Sinn, ein lokales Gebiet zu betrachten für das dann die Infrastruktur dimensioniert wird. Für die Auslegung können die Daten aus dem Bebauungsplan für ein

Neubau- bzw. Transformationsgebiet mit einem ähnlichen Gebiet, für das eine Clusteranalyse bereits durchgeführt wurde, verglichen werden, so dass die Dimensionierung der neuen Infrastruktur auf die zu erwartende Einwohnerzahl erfolgen kann. So können die Ergebnisse dieser Arbeit zur zukünftigen Berechnung von Trinkwasserinfrastrukturen herangezogen werden.

Die angewandte Methode der Datensynthese mit Hilfe der Software FME hat zum einen gezeigt, dass aus verschiedenen Quellen Daten eingelesen werden können. Die Software zeichnet sich dadurch aus, dass sie in der Lage ist, alle verwendeten Formate einzulesen. Auf der anderen Seite hat sich gezeigt, dass, obwohl das Programm in einer 64-bit Version vorliegt, bei den meisten Berechnungen auf einem 4-Kern-System nur ein Prozessor genutzt wird, was unökonomisch ist.

6. Ausblick

Die Automatisierung der Datensynthese und -analyse durch FME ermöglicht es, für weitere Ortschaften und Gebiete die Siedlungsstruktur durch Clusterbildung zu identifizieren. Voraussetzung ist, dass die nötigen Datenquellen vorhanden sind. 3D-Gebäudedaten gibt es in immer mehr Bundesländern. Der Bestandsdatenauszug ist per Gesetz bestimmt und damit auch in weiten Teilen bereits verfügbar. Die Zensus-Daten liegen für die gesamte Bundesrepublik vor. Im Jahr 2021 steht die nächste Volkszählung an. Dann stehen neue Daten zur Verfügung.

Sollen für ein größeres Gebiet oder ein Bundesland die Berechnungen mit FME durchgeführt werden, ist ein wesentlich leistungsfähigerer Rechner nötig. Im Rahmen dieser Arbeit wurde versucht den Hochleistungsrechner der TU Darmstadt, den Lichtenberg High Performance Computer (LHPC), für die Berechnungen zu nutzen. Dies konnte nicht gelingen, da die zur Verfügung gestellte Software in der *Database Edition* von den Entwicklern gegen Bibliotheken kompiliert wurde, die auf dem LHPC nicht oder nur teilweise vorhanden sind. Ein manuelles Installieren dieser Bibliotheken im eigenen Home-Verzeichnis war erfolgreich, doch waren die verwendeten Versionen nach kurzer Zeit durch die regelmäßigen Updates des gesamten Systems durch die Administratoren inkompatibel zu den aktualisierten Bibliotheken des Systems. Es existiert eine FME-Version, die nicht gegen die grafischen Bibliotheken kompiliert wurde, die reine *engine*. Es bleibt zu klären, ob diese in der Lage ist, auf mehreren Kernen bzw. Rechen-Clustern gleichzeitig zu laufen.

Die Ergebnisse dieser Arbeit werden wie bereits beschrieben in eine integrierte Wasserbedarfsprognose im *Innogy* Forschungsvorhaben einfließen. Dadurch wird es möglich sein, „den künftigen Wasserbedarf kleinräumlich, einfach und belastbar abzuschätzen - unabhängig vom ohnehin nicht besser prognostizierbaren individuellen Trinkwasser-Verbrauchsverhalten“, wie in der Aufgabenstellung zur vorliegenden Arbeit gefordert.

Danksagung

Dank geht an:

Die Mitarbeiter des IWW Rhein-Main für die freundliche Unterstützung bei der Erstellung dieser Arbeit

Die Firma SafeSoftware für die kostenfreie Bereitstellung der Software FME

Literatur

- „ALKIS-Objektartenkatalog“ (2008). In: *Dokumentation zur Modellierung der Geoinformationen des amtlichen Vermessungswesens (GeoInfoDok)*. Hrsg. von Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV). Version 6.0. URL: <http://www.adv-online.de/AAA-Modell> (besucht am 16.10.2017).
- Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV), Hrsg. (2013). *Gemeinschaftsportal für den Zugang zum vernetzten Bodenrichtwertinformationssystem (VBORIS) der Gutachterausschüsse für Grundstückswerte in den Ländern der Bundesrepublik Deutschland*. URL: <http://www.adv-online.de/AdV-Produkte/Standards-und-Produktblaetter/Standards-des-Liegenschaftskatasters> (besucht am 09.10.2017).
- AdV, Hrsg. (2016). *Richtlinie über Gebühren für die Bereitstellung und Nutzung von Geobasisdaten der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV-Gebührenrichtlinie - AdV-GR)*. URL: <http://www.adv-online.de/AdV-Produkte/Standards-und-Produktblaetter/Standards-des-Liegenschaftskatasters> (besucht am 09.10.2017).
- Arbeitsgemeinschaft der Vorsitzenden der Gutachterausschüsse für Grundstückswerte in Nordrhein-Westfalen (AGVGA.NRW), Hrsg. (2017). *BORISplus.NRW - Bodenrichtwerte Nordrhein-Westfalen*. Coesfeld. URL: <https://www.boris.nrw.de/borisplus/?lang=de> (besucht am 21.09.2017).
- Albers, Gerd und Julian Wékel (2011). *Stadtplanung: eine illustrierte Einführung*. 2., durchges. und aktualisierte Aufl. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Backhaus, Klaus et al. (2016). *Multivariate Analysemethoden*. 14. Aufl. Springer Gabler Verlag.
- Bellinghoff, Markus, Christian Dahmen und Christian Heisig (2015). *FME Desktop : das deutschsprachige Handbuch für Einsteiger und Anwender*. Hrsg. von con terra GmbH. Berlin, Offenbach: Wichmann-Verlag.
- Berendes, Konrad, Walter Frenz und Hans-Jürgen Müggenborg, Hrsg. (2017). *WHG: Wasserhaushaltsgesetz - Kommentar*. Berlin: Erich Schmidt Verlag.
- Bill, Ralf (2016). *Grundlagen der Geo-Informationssysteme*. 6., völlig neu bearbeitete und erweiterte Auflage. Berlin: Wichmann-Verlag.
- Boeddinghaus, Gerhard (2000). *BauNVO - Baunutzungsverordnung: Kommentar*. 4. Aufl. München.

-
- Statistisches Bundesamt, Hrsg. (2016). *Datenangebot zum Zensusatlas - Klassifizierte Ergebnisse*. Wiesbaden.
- Dieckmann, Andreas (03/2012). *Empirische Sozialforschung*. 6. Auflage. Hamburg: Rowohlt Taschenbuch Verlag.
- „DVGW-W 410“ (12/2008). „DVGW Arbeitsblatt W 410 - Wasserbedarf - Kennwerte und Einflussgrößen“. In: *DVGW-Regelwerk*. Hrsg. von Deutscher Verein des Gas- und Wasserfaches e. V. Bonn.
- Fritsch, Peter et al. (2014). *Taschenbuch der Wasserversorgung*. 16., vollst. überarb. und aktual. Auflage. Wiesbaden: Springer Vieweg.
- Springer Gabler Verlag, Hrsg. (2015). *Gabler Wirtschaftslexikon*. Wiesbaden: Springer Gabler Verlag. URL: <http://wirtschaftslexikon.gabler.de/Archiv/5780/hempel-oppenheim-schema-v5.html> (besucht am 06.09.2017).
- Grossmann, Jörg und Helmut Hofmann (2008). „Integrierte Wasserbedarfsprognose - Teil 1: Erstellung eines innovativen Prognosemodells für Hamburg Wasser“. In: *GWF Wasser/Abwasser* 149/20, S. 758–763.
- Haffner, Yvonne (2004). „Sozialwissenschaftliche Modellierung zur Privatisierung der Wasserversorgung : Aspekte der Nachhaltigkeit“. Diss. Darmstadt: Inst. WAR, S. 764–772.
- Herber, Werner, Holger Wagner und Ulrich Roth (2008). „Der Regionale Wasserbedarfsnachweis der Hessenwasser GmbH & Co. KG“. In: *GWF Wasser/Abwasser* 149/20, S. 773–779.
- Hiessl, Harald (2001). „Wasserbedarf - ein Konzept im Wandel“. In: *Der Wassersektor in Deutschland: Methoden und Erfahrungen*. Hrsg. von Karl-Ulrich Rudolph. Berlin: Umweltbundesamt, S. 44–58.
- Holland, Heinrich und Kurt Scharnbacher (2006). *Grundlagen der Statistik*. 7., aktualisierte Auflage. Wiesbaden: Betriebswirtschaftlicher Verlag Dr. Th. Gabler.
- Karger, Rosemarie, Claus Cord-Landwehr und Frank Hoffmann (2008). *Wasserversorgung*. 13. Aufl. Wiesbaden: Vieweg + Teubner Verlag.
- al-Khouri, Giyas (2003). „Automatische Erkennung von Gebäudetypen mit Hilfe wissensbasierter Methoden unter Einbeziehung funktionsbezogener und kontextueller Informationen“. Diss. Dresden: Technische Universität Dresden.

-
- Kluge, Thomas, Jens Libbe und Engelbert Schramm (2006). „Auswirkungen des demographischen Wandels auf Trinkwasserversorgung und Abwasserbeseitigung“. In: *Praktische Konzepte zur demographischen Stadtentwicklung*. Hrsg. von Jürgen Dettbarn-Reggentin und Heike Reggentin.
- Kluge, Thomas, Jutta Deffner et al. (2008). „Integrierte Wasserbedarfsprognose - Teil 2: Grundlagen und Methodik“. In: *GWF Wasser/Abwasser* 149/20, S. 764–772.
- Koegst, Thilo und Jens Tränckner (2010). „Demographische und klimatische Auswirkungen auf Trinkwasserverbrauch, Schmutzwasseranfall und dessen Zusammensetzung“. In: *Anpassung der Abwassersysteme an veränderte Randbedingungen*. Hrsg. von Peter Krebs. Bd. 33. Dresden: Inst. für Siedlungs- und Industrierwasserwirtschaft.
- Korda, Martin, Hrsg. (2005). *Städtebau : technische Grundlagen*. 5., neubearb. Aufl. Stuttgart.
- Kühnel, Steffen-M. und Dagmar Krebs (07/2007). *Statistik für die Sozialwissenschaften*. 4. Auflage. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag.
- Malz, Simone und Ulrich Scheele (2011). „Wasserbedarf und Wasserverbrauch privater Haushalte und der Industrie nach Ländern“. In: *WARNSIGNAL KLIMA: Genug Wasser für alle?* 3. Ausgabe. Hrsg. von José L. Lozán. URL: http://www.climate-service-center.de/imperia/md/content/csc/warnsignalklima/warnsignal_klima_wasser_kap2_2.2_malz_scheele.pdf (besucht am 21.08.2017).
- Martin, Tobias, Andreas Korth und Konrad Schubert (2017). *Aktualisierung der Verbrauchsganglinien für Haushalte, Kleingewerbe und öffentliche Gebäude sowie Entwicklung eines Modells zur Simulation des Wasserbedarfs*. DVGW Deutscher Verein des Gas- und Wasserfaches e. V.
- Masanek, Rosemarie (1996). „Technische und soziale Einflüsse auf die Entwicklung des Trinkwasserbedarfs“. Diss. Hannover: Universität Hannover, S. 764–772.
- Merbecks, Stefanie et al. (2015). *Baunetz Wissen - Nachhaltig Bauen*. Hrsg. von BauNetz Media GmbH. Berlin. URL: http://www.baunetzwissen.de/standardartikel/Nachhaltig-Bauen_Gebaeudeform_662875.html (besucht am 05.10.2017).
- Neunteufel, Roman, Laurent Richard, Reinhard Perfler et al. (01/2012). *Wasserverbrauch und Wasserbedarf - Auswertung empirischer Daten zum Wasserverbrauch*. Hrsg. von Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft Österreich. Wien. URL: http://www.bmlfuw.gv.at/publikationen/wasser/wasserwirtschaft_wasserpolitik/wasserverbrauch_und_wasserbedarf.html (besucht am 03.11.2015).

-
- QGIS Project, Hrsg. (2017). *QGIS Benutzerhandbuch*. URL: <http://docs.qgis.org/2.18/pdf/de/QGIS-2.18-UserGuide-de.pdf> (besucht am 12.10.2017).
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Wien. URL: <https://www.R-project.org> (besucht am 07.12.2017).
- Renzetti, Steven (2005). „The Economics of Water Demands“. In: *Natural Resource Management and Policy* 22, S. 145–156.
- Ruf, Hans-Ulrich (2016). *Bildkommentar DIN 276/DIN 277 : Kosten im Bauwesen, Grundflächen und Rauminhalte im Bauwesen*. Stuttgart: BKI GmbH Baukosteninformationszentrum Deutscher Architektenkammern.
- Schleich, Joachim und Thomas Hillenbrand (04/2007). „Determinants of residential water demand in Germany“. In: *Ecological Economics* 68.6, S. 1756–1769. URL: http://www.sciencedirect.com/science/article/pii/S0921800908004977?via%3Dihub;%20http://acels-cdn.com/S0921800908004977/1-s2.0-S0921800908004977-main.pdf?_tid=307314c6-75eb-11e7-8cde-00000aabb0f01&acdnat=1501504046_beed309659c3b19cb68e30cf71c884b;%20https://idw-online.de/de/news211686;%20http://www.isi.fhg.de/publ/downloads/isi07a05/residential-water-demand-in-germany.pdf (besucht am 21.08.2017).
- Schmitz, Gottfried (2004). *Anpassung raumplanerischer Instrumente an Veränderungen der Bevölkerungsentwicklung*. Bd. 3. Räumliche Kosequenzen des demograpischen Wandels. Hannover: Akademie für Raumforschung und Landesplanung, S. 127–143.
- Schumann, Daniel und Andreas Korth (2014). „Ergebnisse aktueller Untersuchungen zum Wasserverbrauchsverhalten“. In: *Neue Methoden und Verfahren für die Wasserversorgung*. Hrsg. von DVGW-Technologiezentrum Wasser. Bd. 19. Karlsruhe: DVGW-Technologiezentrum Wasser.
- Walker, T. C. und R. K. Miller (1990). „Geographic Information System. An Assessment of Technology, Applications and Products“. In: *SEAI Technical Publications* 1.
- Winje, Dietmar und Josef Iglhaut (1983). *Wasserbedarfsprognose : Der Wasserbedarf in der Bundesrepublik Deutschland bis zum Jahre 2010*. 149/20. Berlin: E. Schmidt, S. 764–772.
- Wricke, Burkhard (2005). „Technische Rahmenbedingungen für den Stadtumbau - Bewertung der Auswirkungen auf die Wasserverteilung.“ In: *gwf - Wasser|Abwasser* 146.13, S. 31–33. URL: http://www.tzw.de/de/abteilungen/aussenstelle_dresden/publikationen_2005-578 (besucht am 04.09.2017).

Zimmermann, Peter (2010). *Immobilienwertermittlungsverordnung: Kommentar*. München: Beck-Verlag.

A. Ergebnisse der Datenaufbereitung

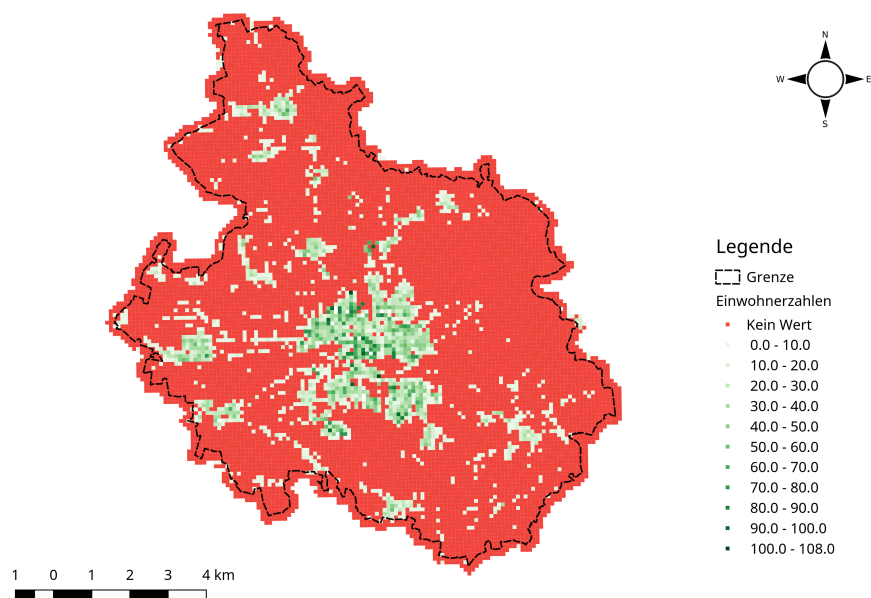


Abbildung A.1: Ergebnis Z0 Reduktion der Zensus-Daten auf Zielgebiet, hier: Stadt Lemgo

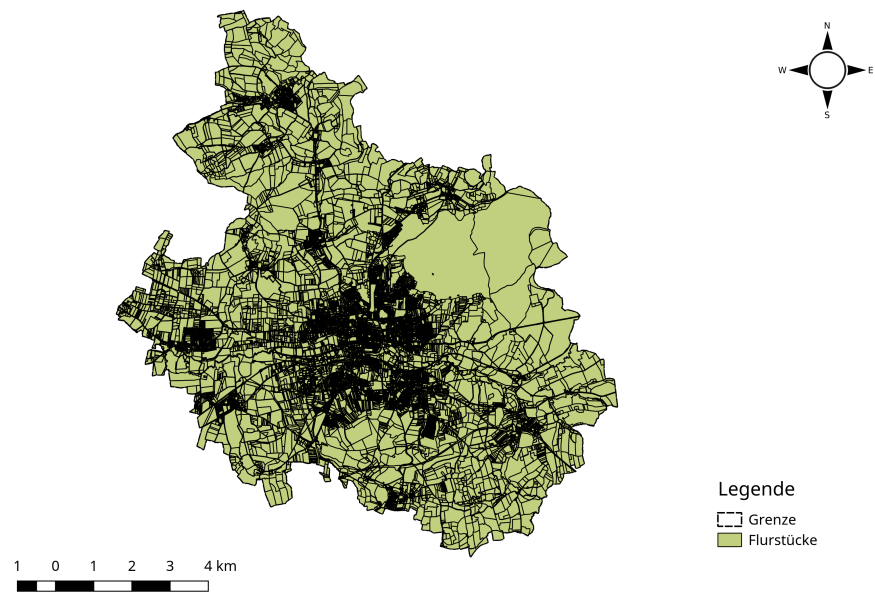


Abbildung A.2: Ergebnis *F0* Clippen der Flurstücke, hier: Stadt Lemgo

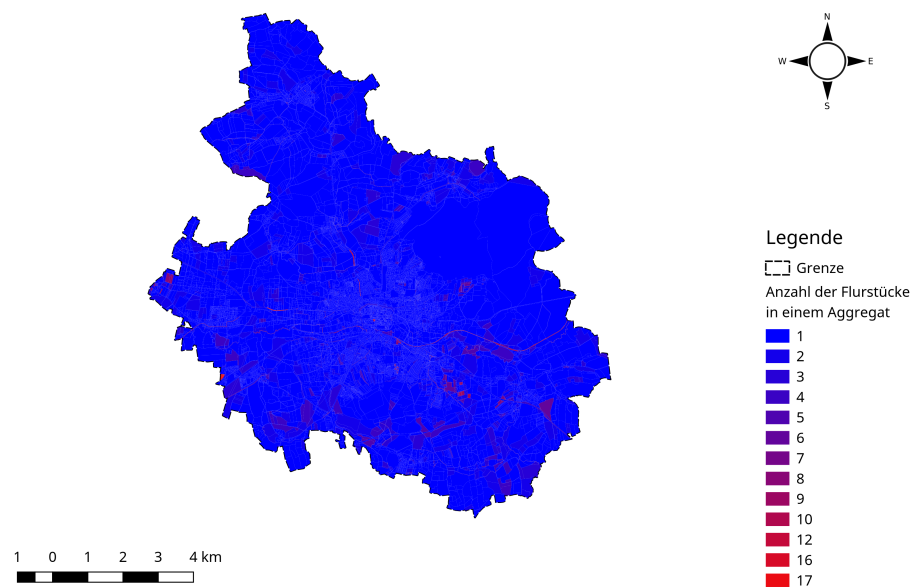


Abbildung A.3: Ergebnis *F1* Aggregation der Flurstücke, hier: Stadt Lemgo

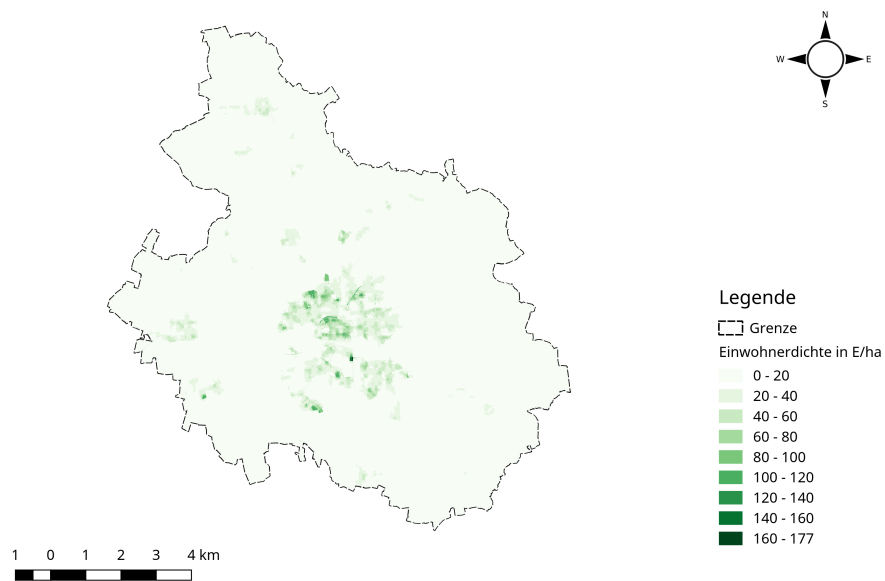


Abbildung A.4: Ergebnis *F2* Projizieren der Zensus-Daten auf Flurstücke, hier: Stadt Lemgo

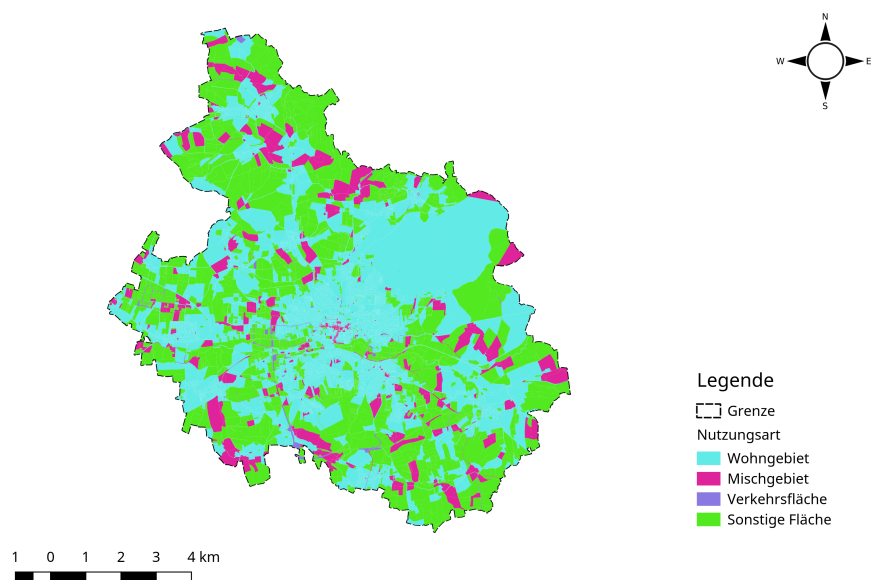


Abbildung A.5: Ergebnis *F3* Verheiraten Flurstücke und Nutzungsart, hier: Stadt Lemgo

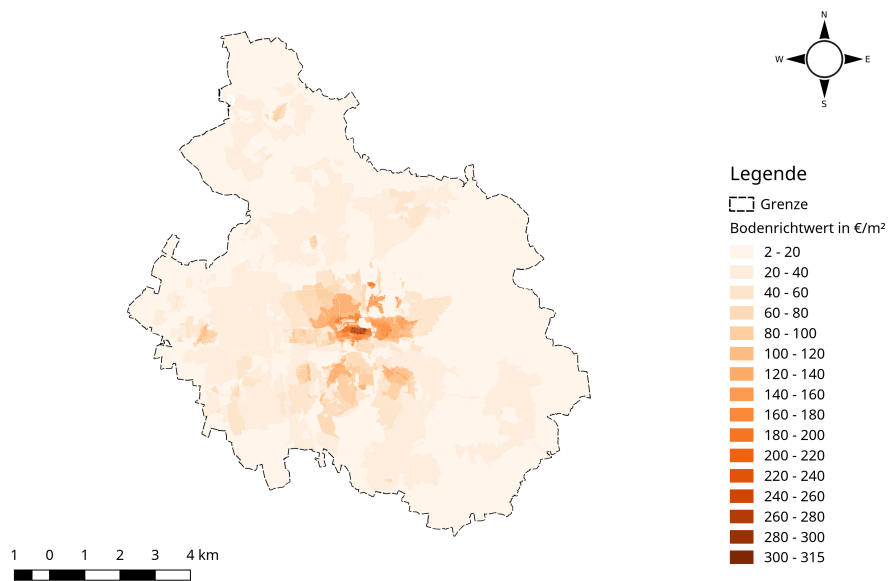


Abbildung A.6: Ergebnis *F4* Projizieren der Bodenrichtwerte auf Flurstücke, hier: Stadt Lemgo

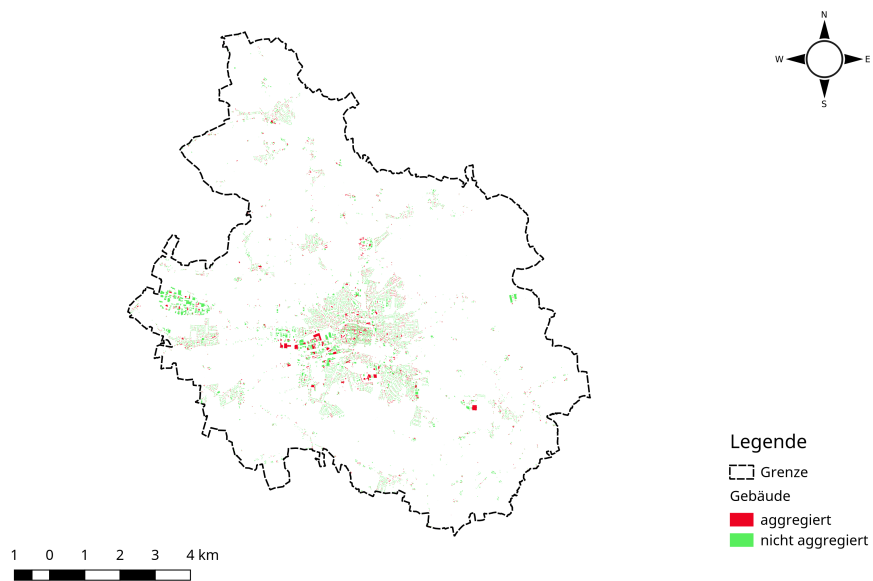


Abbildung A.7: Ergebnis *G0* Clippen der 3D-Gebäude und Reduktion der XML-Ebenen, hier: Stadt Lemgo

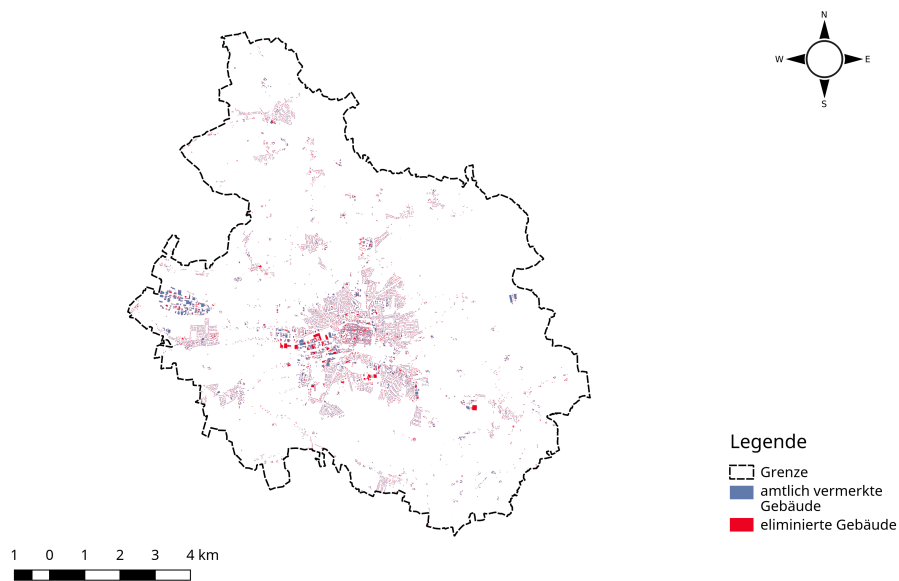


Abbildung A.8: Ergebnis G2 Verheiraten von Gebäudedaten, hier: Stadt Lemgo

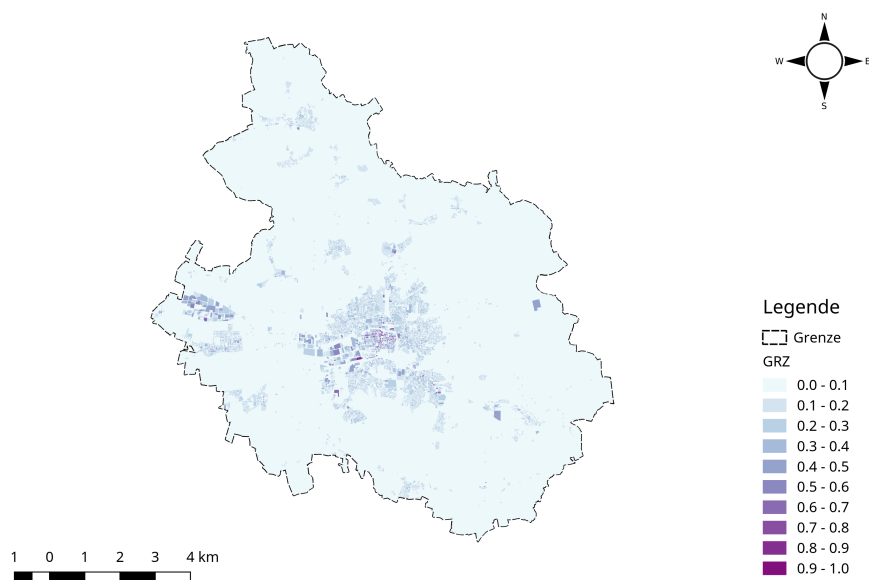


Abbildung A.9: Ergebnis GF Verheiraten der Daten - Grundflächenzahl, hier: Stadt Lemgo

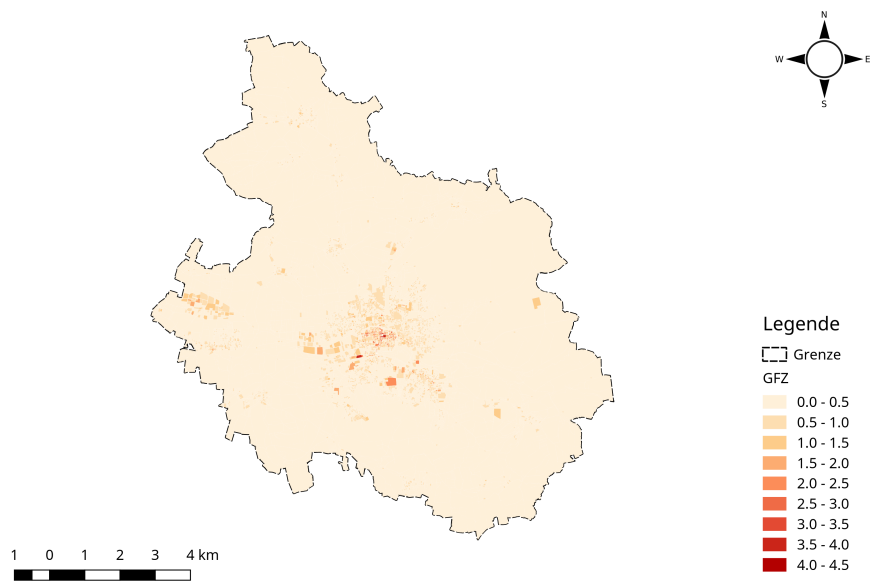


Abbildung A.10: Ergebnis *GF Verheiraten der Daten* - Geschossflächenzahl, hier: Stadt Lemgo

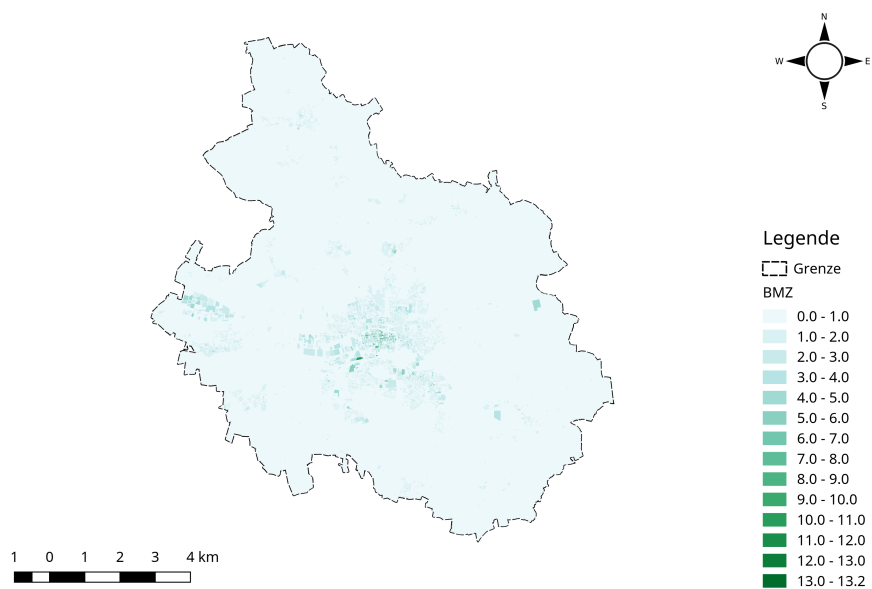


Abbildung A.11: Ergebnis *GF Verheiraten der Daten* - Baumassenzahl, hier: Stadt Lemgo

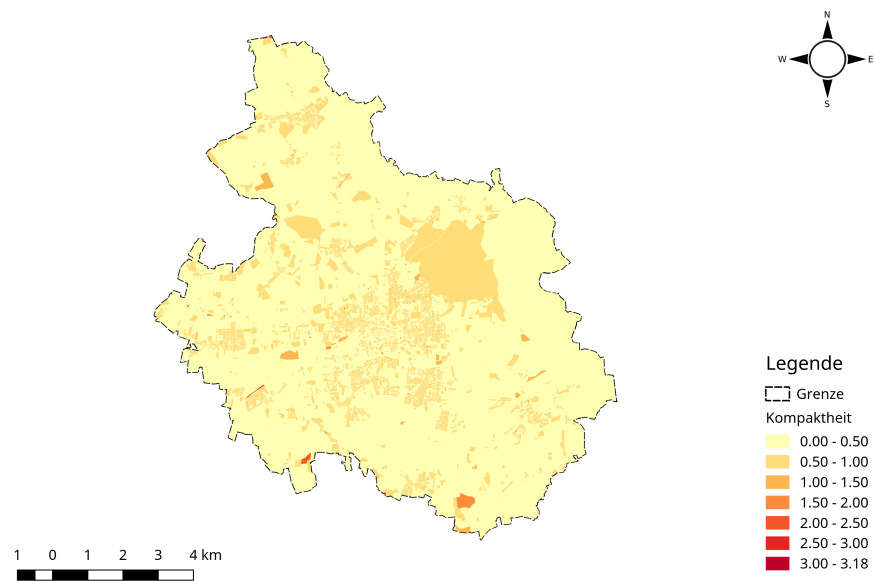


Abbildung A.12: Ergebnis GF Verheiraten der Daten - Kompaktheit, hier: Stadt Lemgo

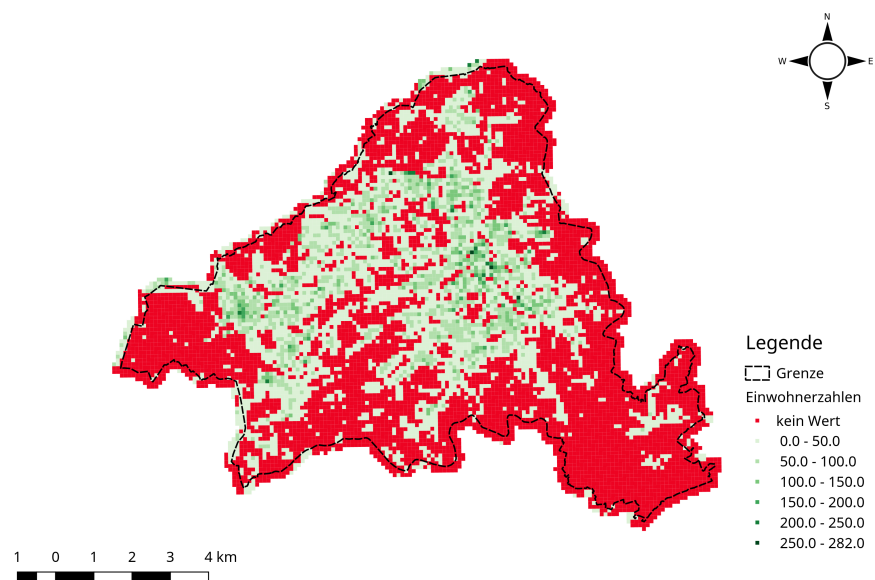


Abbildung A.13: Ergebnis Z0 Reduktion der Zensus-Daten auf Zielgebiet, hier: Stadt Solingen

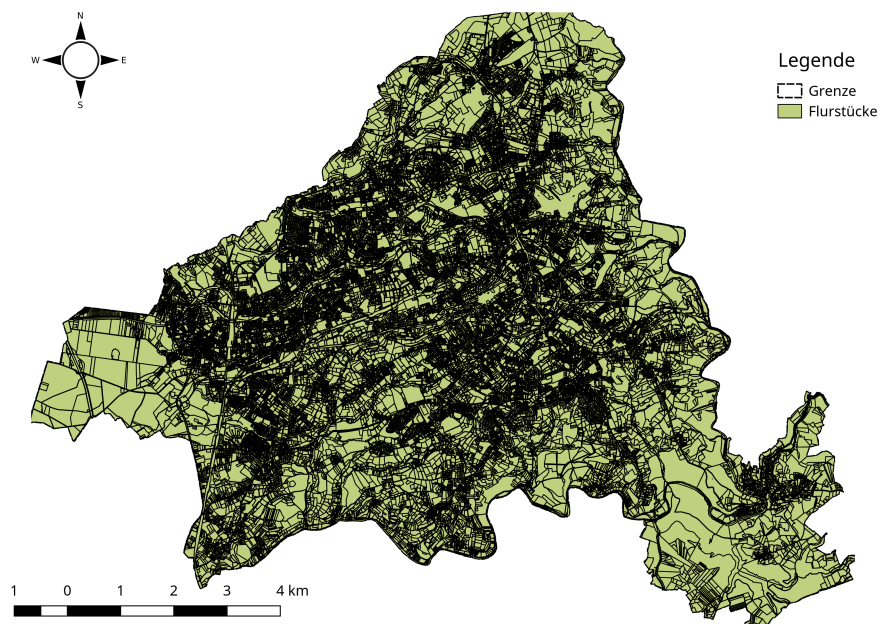


Abbildung A.14: Ergebnis *F0* Clippen der Flurstücke, hier: Stadt Solingen

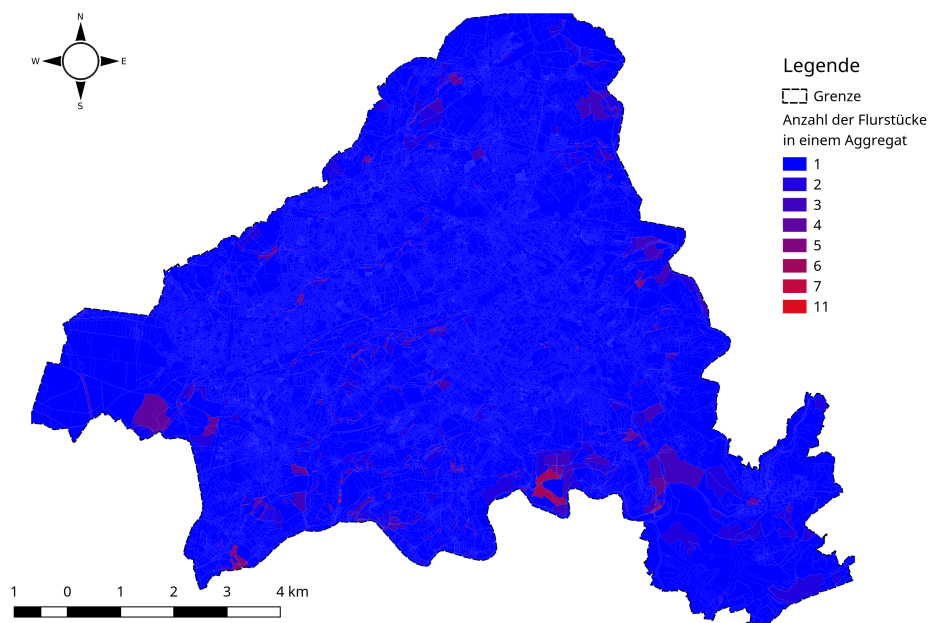


Abbildung A.15: Ergebnis *F1* Aggregation der Flurstücke, hier: Stadt Solingen

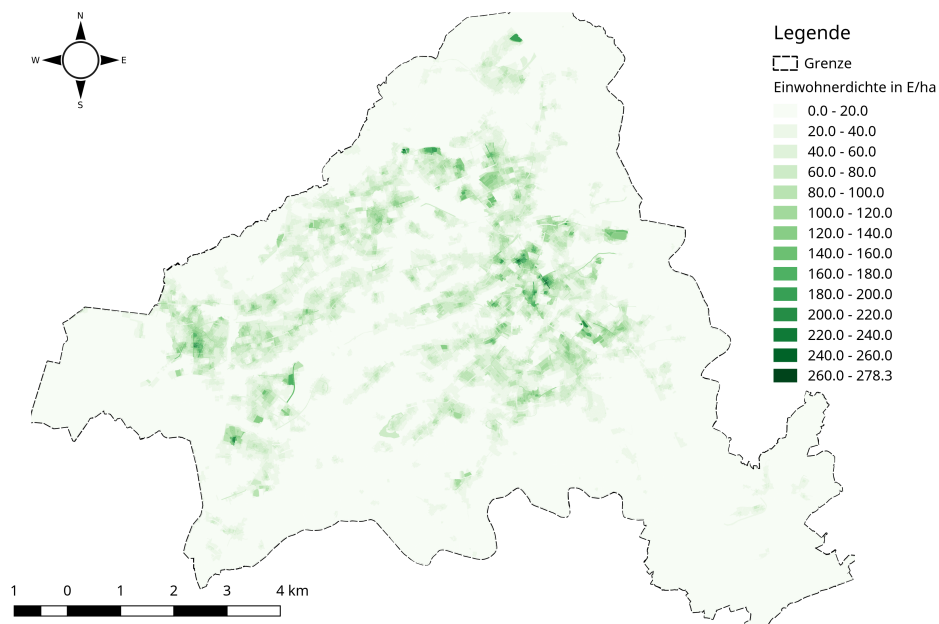


Abbildung A.16: Ergebnis F2 Projizieren der Zensus-Daten auf Flurstücke, hier: Stadt Solingen

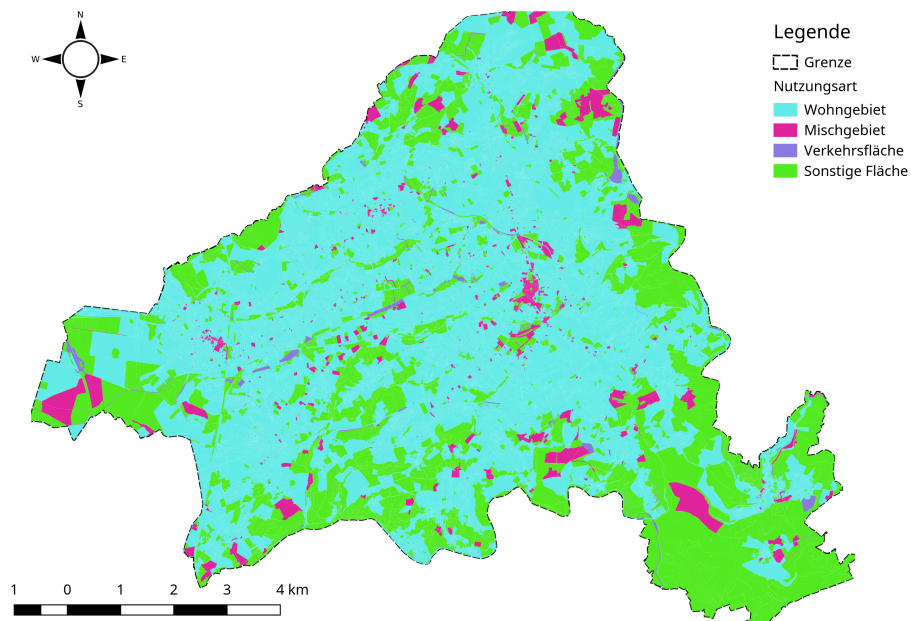


Abbildung A.17: Ergebnis F3 Verheiraten Flurstücke und Nutzungsart, hier: Stadt Solingen

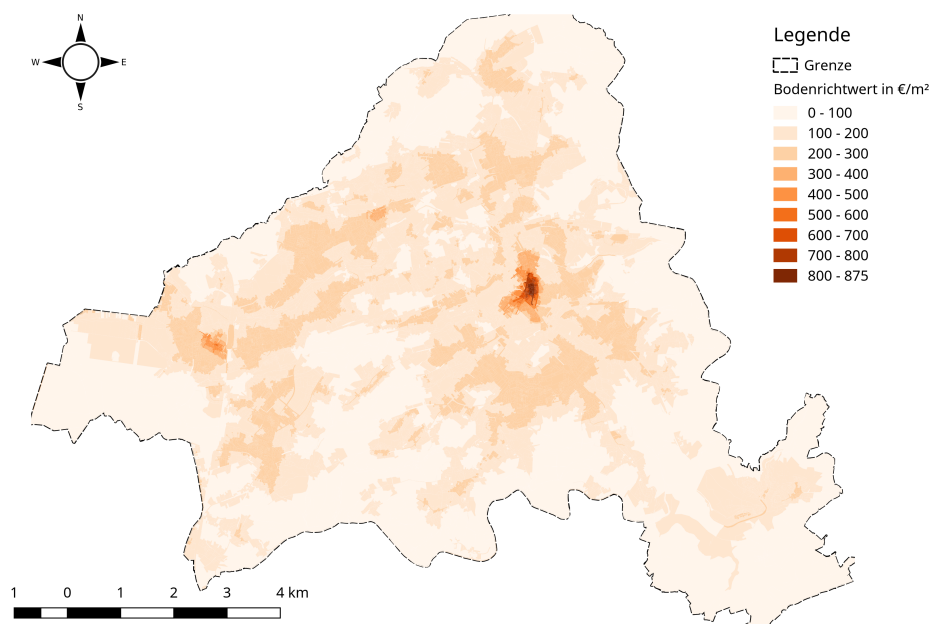


Abbildung A.18: Ergebnis *F4* Projizieren der Bodenrichtwerte auf Flurstücke, hier: Stadt Solingen

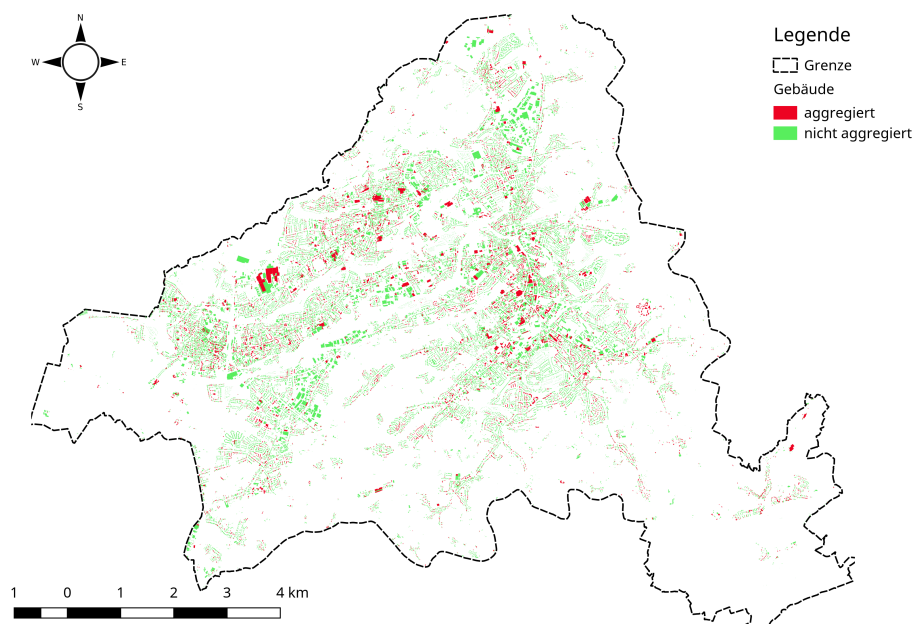


Abbildung A.19: Ergebnis *G0* Clippen der 3D-Gebäude und Reduktion der XML-Ebenen, hier: Stadt Solingen

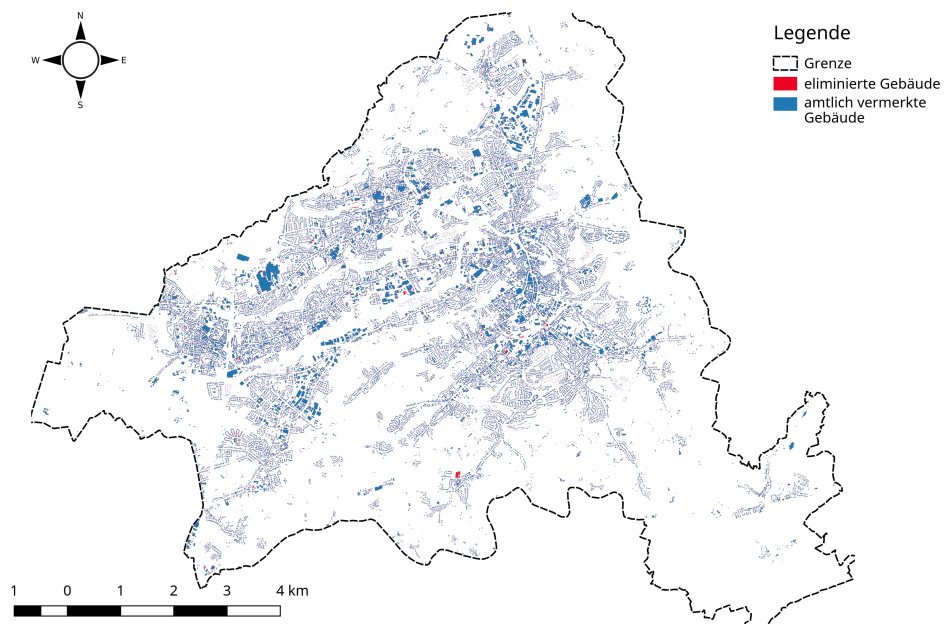


Abbildung A.20: Ergebnis G2 Verheiraten von Gebäudedaten, hier: Stadt Solingen

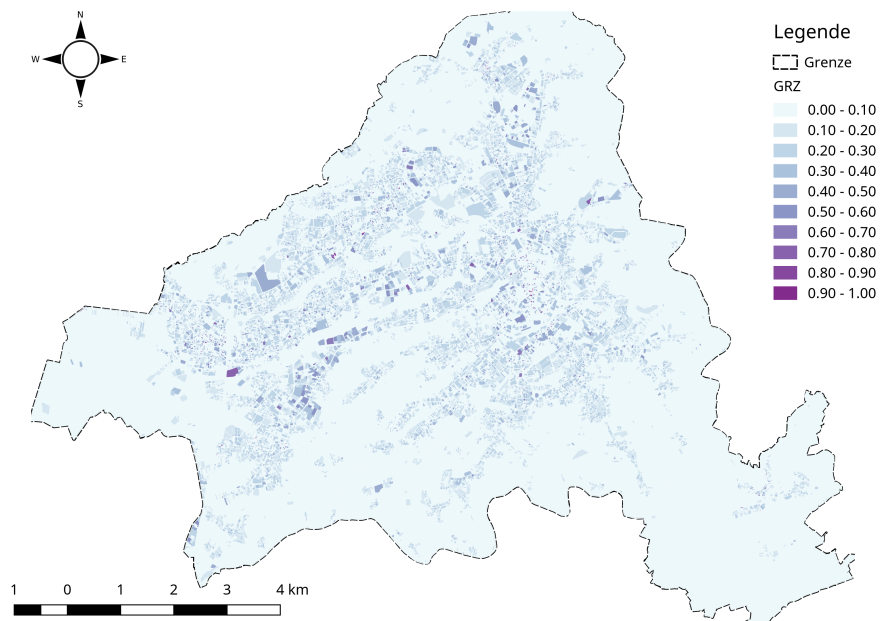


Abbildung A.21: Ergebnis GF Verheiraten der Daten - Grundflächenzahl, hier: Stadt Solingen

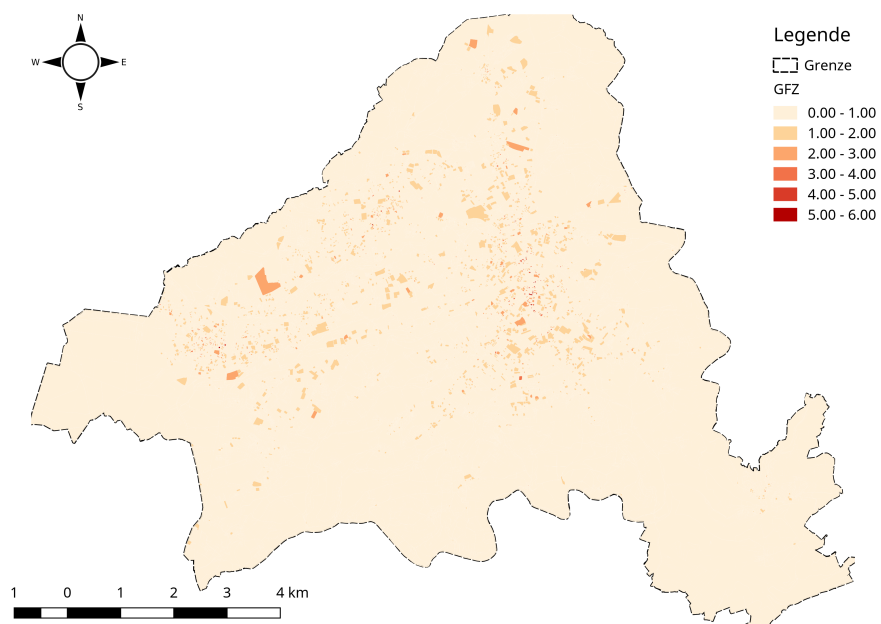


Abbildung A.22: Ergebnis *GF Verheiraten der Daten* - Geschossflächenzahl, hier: Stadt Solingen

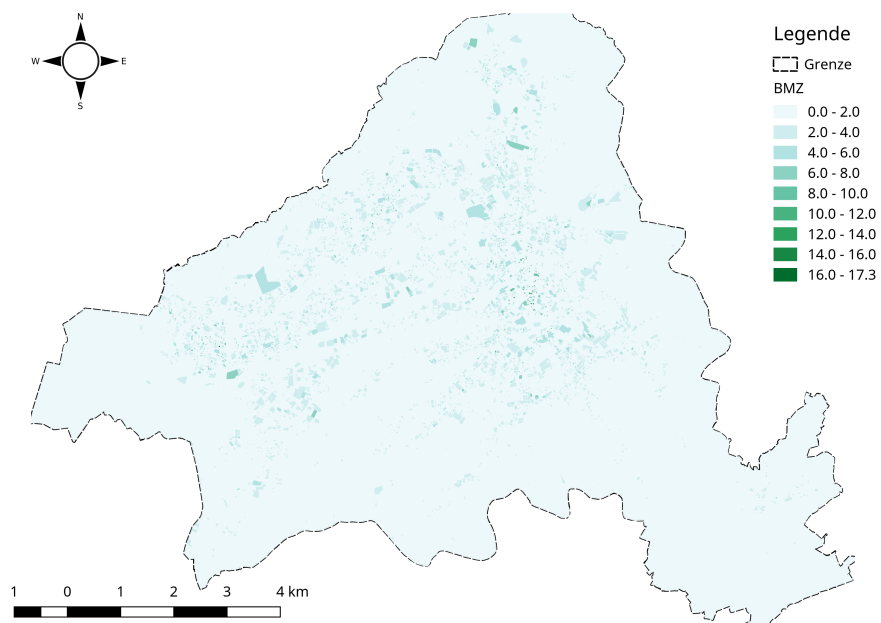


Abbildung A.23: Ergebnis *GF Verheiraten der Daten* - Baumassenzahl, hier: Stadt Solingen

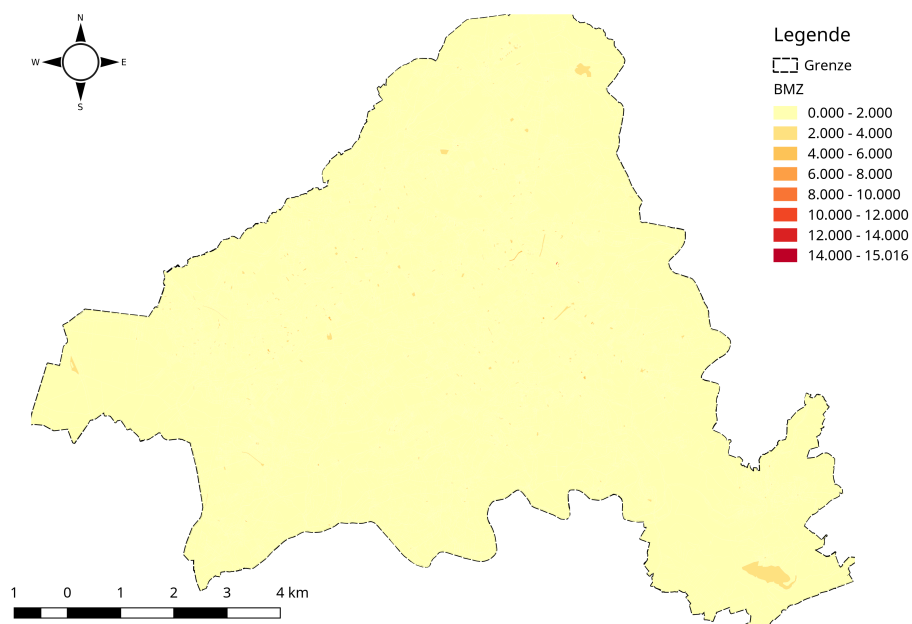


Abbildung A.24: Ergebnis *GF* Verheiraten der Daten - Kompaktheit, hier: Stadt Solingen

B. Ergebnisse der Datenanalyse

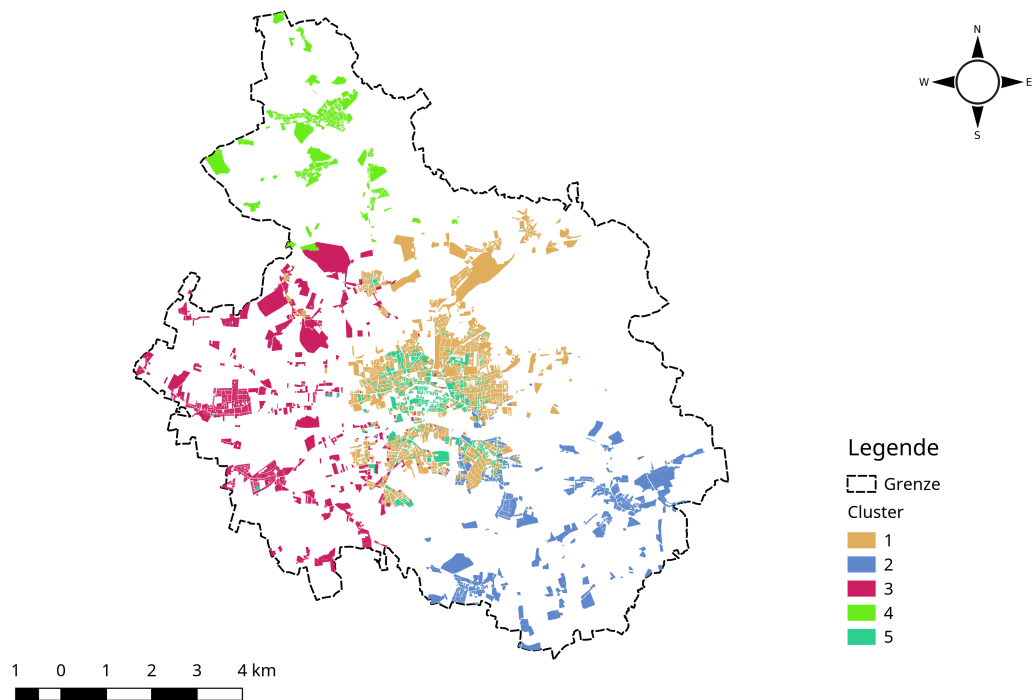
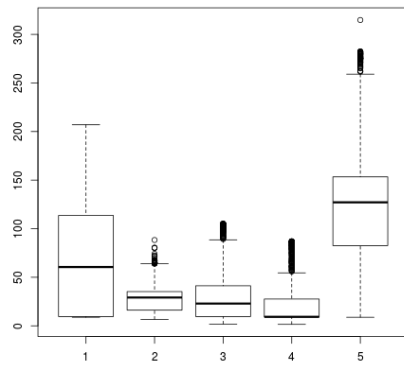
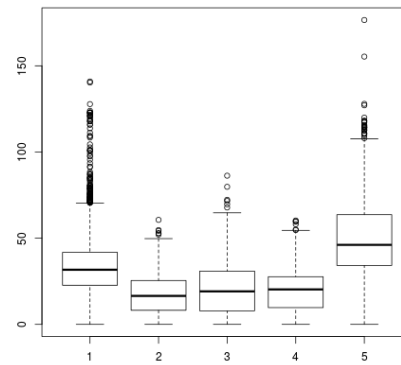


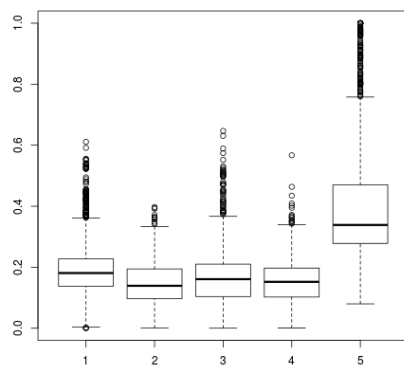
Abbildung B.1: Ergebnis GF Clusteranalyse, hier: Stadt Lemgo



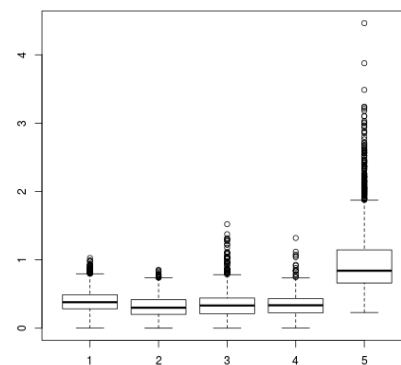
(a) Bodenrichtwert



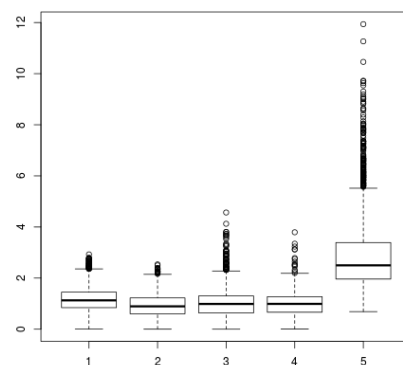
(b) Einwohnerdichte



(c) Grundflächenzahl



(d) Geschossflächenzahl



(e) Baummassenzahl

Abbildung B.2: Boxplots zur Visualisierung der Ergebnisse, hier: Stadt Lemgo

Cluster	Variable	0%-Quartil	25%-Quartil	50%-Quartil	75%-Quartil	100%-Quartil
1	Bodenrichtwert	8.800000	9.566667	60.500000	113.582812	206.951496
	Einwohnerdichte	0.00000	22.69050	31.67976	41.78242	140.89857
	Grundflächenzahl	0.0007846776	0.1376068974	0.1808258578	0.2273269744	0.6109564654
	Geschossflächenzahl	0.0008900701	0.2837746812	0.3800541572	0.4889051361	1.0270737602
	Baummassenzahl	0.00267021	0.84246092	1.12809568	1.44956105	2.92759894
2	Bodenrichtwert	6.70000	16.13333	29.13333	35.23333	88.33333
	Einwohnerdichte	0.000000	8.185235	16.508772	25.451845	60.609780
	Grundflächenzahl	0.0004393774	0.0969519183	0.1391187656	0.1944347813	0.3972824471
	Geschossflächenzahl	0.0004020304	0.2022243440	0.3000898853	0.4163961091	0.8529233808
	Baummassenzahl	0.001206091	0.601204549	0.894211098	1.222511093	2.536706741
3	Bodenrichtwert	1.833333	9.566667	22.900000	41.233333	105.000000
	Einwohnerdichte	0.000000	7.889754	19.141523	30.912164	86.251928
	Grundflächenzahl	0.0001298048	0.1038432787	0.1608464676	0.2101414902	0.6468639289
	Geschossflächenzahl	0.0001052752	0.2126559617	0.3318119691	0.4409070490	1.5222738365
	Baummassenzahl	0.0003158255	0.6354484626	0.9871524426	1.3025661647	4.5668215094
4	Bodenrichtwert	1.500000	9.233333	9.233333	27.566667	87.015726
	Einwohnerdichte	0.000000	9.722536	20.267131	27.662874	60.248720
	Grundflächenzahl	0.000809139	0.102218055	0.152033466	0.197232407	0.566886976
	Geschossflächenzahl	0.001188895	0.225354811	0.335348261	0.433034974	1.319912158
	Baummassenzahl	0.003566685	0.663731758	0.990397960	1.274330060	3.793382007
5	Bodenrichtwert	8.80000	82.58798	127.09521	153.34159	314.78118
	Einwohnerdichte	0.00000	34.20893	46.12106	63.64053	176.56759
	Grundflächenzahl	0.0798171	0.2780028	0.3385616	0.4701908	1.0000579
	Geschossflächenzahl	0.2292926	0.6607287	0.8410705	1.1457745	4.4632075
	Baummassenzahl	0.6878779	1.9625371	2.4994138	3.3896541	11.9367870

Tabelle B.1: Statistische Kennzahlen der Cluster, hier: Stadt Lemgo

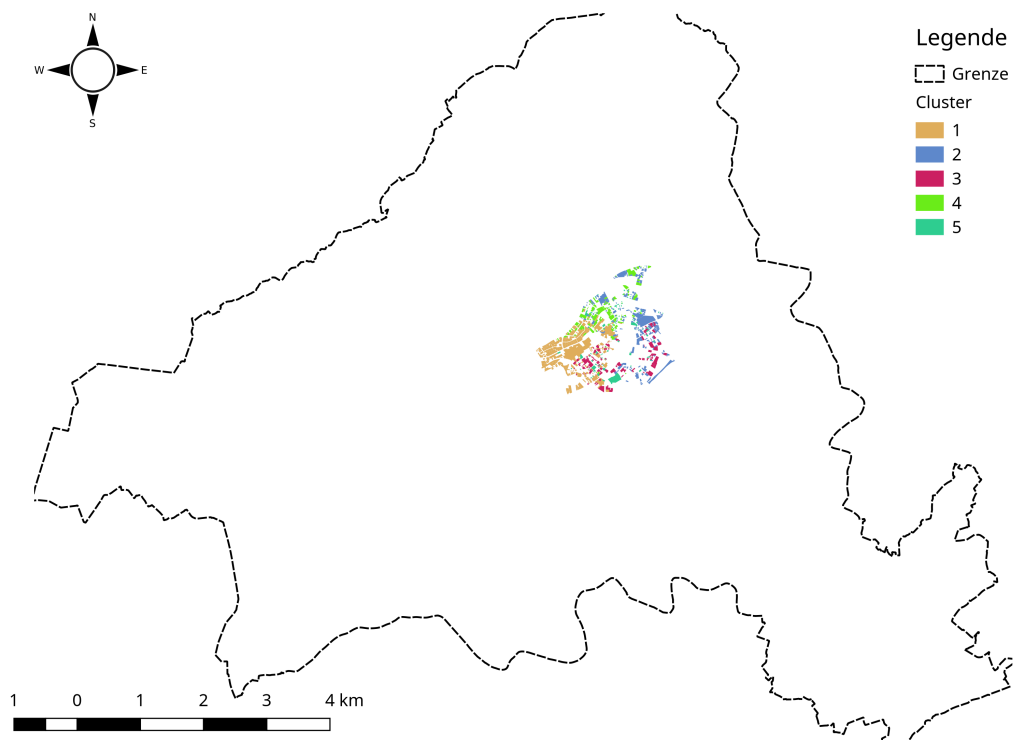
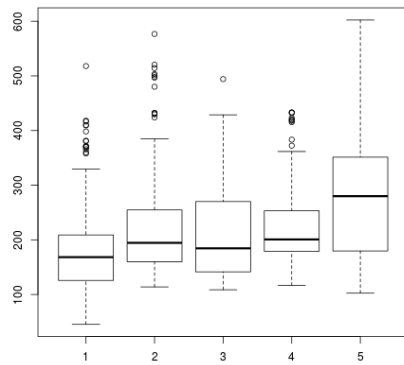
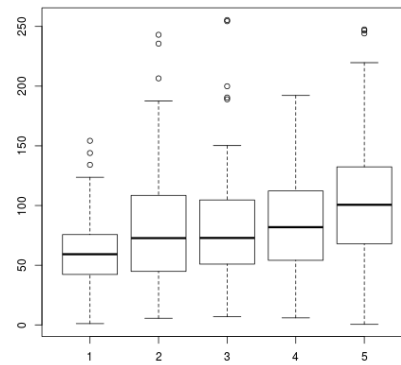


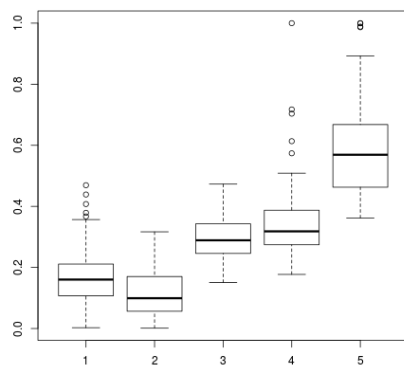
Abbildung B.3: Ergebnis *GF* Clusteranalyse, hier: Solingen Mitte



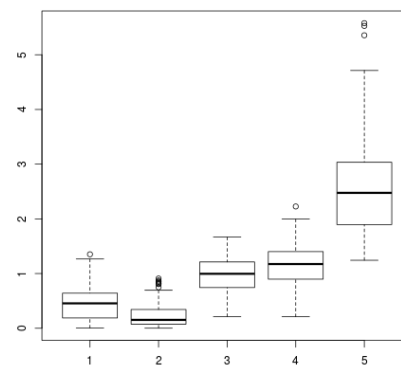
(a) Bodenrichtwert



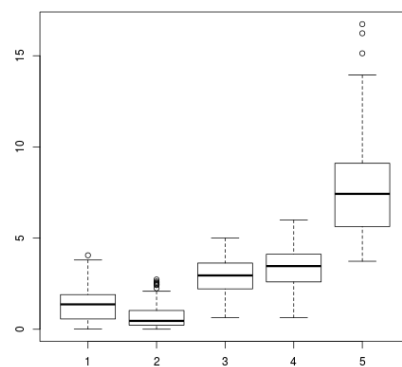
(b) Einwohnerdichte



(c) Grundflächenzahl



(d) Geschossflächenzahl



(e) Baumassenzahl

Abbildung B.4: Boxplots zur Visualisierung der Ergebnisse, hier: Solingen Mitte

Cluster	Variable	0%-Quartil	25%-Quartil	50%-Quartil	75%-Quartil	100%-Quartil
1	Bodenrichtwert	45.34529	125.74475	168.31038	208.96098	517.93940
	Einwohnerdichte	1.097949	42.263317	59.226756	75.676582	154.227281
	Grundflächenzahl	0.002713595	0.107108792	0.160254093	0.211052535	0.469344329
	Geschossflächenzahl	0.003145622	0.187497630	0.454981537	0.641176432	1.352123879
	Baummassenzahl	0.009436865	0.562492891	1.364944612	1.899152911	4.056371636
2	Bodenrichtwert	113.8193	160.0671	194.4831	254.7511	576.6954
	Einwohnerdichte	5.55532	44.79946	72.70501	108.52359	243.02958
	Grundflächenzahl	0.001081704	0.056499270	0.099279087	0.170227566	0.316690430
	Geschossflächenzahl	0.0009197284	0.0740094725	0.1526369515	0.3423572026	0.9088711693
	Baummassenzahl	0.002759185	0.222028417	0.457910855	1.027071608	2.726613508
3	Bodenrichtwert	108.6542	141.6689	184.4605	267.5432	494.1068
	Einwohnerdichte	6.90712	50.98428	72.81851	104.40440	255.39315
	Grundflächenzahl	0.1505643	0.2462181	0.2889056	0.3426874	0.4728923
	Geschossflächenzahl	0.2133065	0.7441111	0.9954991	1.2116282	1.6657679
	Baummassenzahl	0.6399196	2.2200591	2.9505040	3.6141098	4.9973038
4	Bodenrichtwert	116.4883	179.4003	200.7228	250.4706	433.1594
	Einwohnerdichte	5.89565	54.54968	81.88327	111.96929	192.18469
	Grundflächenzahl	0.1772002	0.2741909	0.3181568	0.3864619	1.0000000
	Geschossflächenzahl	0.2095539	0.9024099	1.1735011	1.3994344	2.2260591
	Baummassenzahl	0.6286618	2.6358449	3.4634681	4.1037467	5.9862900
5	Bodenrichtwert	102.4244	179.6529	279.9702	351.2797	602.1953
	Einwohnerdichte	0.4575812	67.9863245	100.6334747	132.3810497	247.4817163
	Grundflächenzahl	0.3617466	0.4627231	0.5691003	0.6679817	1.0000000
	Geschossflächenzahl	1.242703	1.895483	2.475178	3.034327	5.577851
	Baummassenzahl	3.728109	5.632366	7.425533	9.102982	16.733553

Tabelle B.2: Statistische Kennzahlen der Cluster, hier: Solingen Mitte